NUCLEIC ACIDS AND PROTEINS FROM STREPTOCOCCUS GROUPS A & B

All documents cited herein are incorporated by reference in their entirety.

TECHNICAL FIELD

This invention relates to nucleic acid and proteins from the bacteria *Streptococcus agalactiae* (GBS) and *Streptococcus pyogenes* (GAS).

BACKGROUND ART

Once thought to infect only cows, the Gram-positive bacterium *Streptococcus agalactiae* (or "group B streptococcus", abbreviated to "GBS") is now known to cause serious disease, bacteremia and meningitis, in immunocompromised individuals and in neonates. There are two types of neonatal infection. The first (early onset, usually within 5 days of birth) is manifested by bacteremia and pneumonia. It is contracted vertically as a baby passes through the birth canal. GBS colonises the vagina of about 25% of young women, and approximately 1% of infants born via a vaginal birth to colonised mothers will become infected. Mortality is between 50-70%. The second is a meningitis that occurs 10 to 60 days after birth. If pregnant women are vaccinated with type III capsule so that the infants are passively immunised, the incidence of the late onset meningitis is reduced but is not entirely eliminated.

The "B" in "GBS" refers to the Lancefield classification, which is based on the antigenicity of a carbohydrate which is soluble in dilute acid and called the C carbohydrate. Lancefield identified 13 types of C carbohydrate, designated A to O, that could be serologically differentiated. The organisms that most commonly infect humans are found in groups A, B, D, and G. Within group B, strains can be divided into 8 serotypes (Ia, Ib, Ia/c, II, III, IV, V, and VI) based on the structure of their polysaccharide capsule.

Group A streptococcus ("GAS", *S.pyogenes*) is a frequent human pathogen, estimated to be present in between 5-15% of normal individuals without signs of disease. When host defences are compromised, or when the organism is able to exert its virulence, or when it is introduced to vulnerable tissues or hosts, however, an acute infection occurs. Diseases include puerperal fever, scarlet fever, erysipelas, pharyngitis, impetigo, necrotising fasciitis, myositis and streptococcal toxic shock syndrome.

S.pyogenes is typically treated using antibiotics. Although *S.agalactiae* is inhibited by antibiotics, however, it is not killed by penicillin as easily as GAS. Prophylactic vaccination is thus preferable.

Current GBS vaccines are based on polysaccharide antigens, although these suffer from poor immunogenicity. Anti-idiotypic approaches have also been used (*e.g.* WO99/54457). There remains a need, however, for effective adult vaccines against *S.agalactiae* infection. There also remains a need for vaccines against *S.pyogenes* infection.

It is an object of the invention to provide proteins which can be used in the development of such vaccines. The proteins may also be useful for diagnostic purposes, and as targets for antibiotics.

DISCLOSURE OF THE INVENTION

The invention provides proteins comprising the *S.agalactiae* amino acid sequences disclosed in the examples, and proteins comprising the *S.pyogenes* amino acid sequences disclosed in the examples. These amino acid sequences are the even SEQ IDs between 1 and 10960.

It also provides proteins comprising amino acid sequences having sequence identity to the *S.agalactiae* amino acid sequences disclosed in the examples, and proteins comprising amino acid sequences having sequence identity to the *S.pyogenes* amino acid sequences disclosed in the examples. Depending on the particular sequence, the degree of sequence identity is preferably greater than 50% (*e.g.* 60%, 70%, 80%, 90%, 95%, 99% or more). These proteins include homologs, orthologs, allelic variants and functional mutants. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence. Identity between proteins is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH program (Oxford Molecular), using an affine gap search with parameters *gap open penalty=12* and *gap extension penalty=1*.

Preferred proteins of the invention are GBS1 to GBS689 (see Table IV).

The invention further provides proteins comprising fragments of the *S.agalactiae* amino acid sequences disclosed in the examples, and proteins comprising fragments of the *S.pyogenes* amino acid sequences disclosed in the examples. The fragments should comprise at least *n* consecutive amino acids from the sequences and, depending on the particular sequence, *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 30, 40, 50, 60, 70, 80, 90, 100 or more). Preferably the fragments comprise one or more epitopes from the sequence. Other preferred fragments are (a) the N-terminal signal peptides of the proteins disclosed in the examples, (b) the proteins disclosed in the examples, but without their N-terminal signal peptides, (c) fragments common to the related GAS and GBS proteins disclosed in the examples, and (d) the proteins disclosed in the examples, but without their N-terminal amino acid residue.

The proteins of the invention can, of course, be prepared by various means (*e.g.* recombinant expression, purification from GAS or GBS, chemical synthesis *etc.*) and in various forms (*e.g.* native, fusions, glycosylated, non-glycosylated *etc.*). They are preferably prepared in substantially pure form (*i.e.* substantially free from other streptococcal or host cell proteins) or substantially isolated form. Proteins of the invention are preferably streptococcal proteins.

According to a further aspect, the invention provides antibodies which bind to these proteins. These may be polyclonal or monoclonal and may be produced by any suitable means (*e.g.* by recombinant expression). To increase compatibility with the human immune system, the antibodies may be chimeric or humanised (*e.g.* Breedveld (2000) *Lancet* 355(9205):735-740; Gorman & Clark (1990) *Semin. Immunol.* 2:457-466), or fully human antibodies may be used. The antibodies may include a detectable label (*e.g.* for diagnostic assays).

According to a further aspect, the invention provides nucleic acid comprising the *S.agalactiae* nucleotide sequences disclosed in the examples, and nucleic acid comprising the *S.pyogenes* nucleotide sequences disclosed in the examples. These nucleic acid sequences are the odd SEQ IDs between 1 and 10966.

In addition, the invention provides nucleic acid comprising nucleotide sequences having sequence identity to the *S.agalactiae* nucleotide sequences disclosed in the examples, and nucleic acid comprising nucleotide sequences having sequence identity to the *S.pyogenes* nucleotide sequences disclosed in the examples. Identity between sequences is preferably determined by the Smith-Waterman homology search algorithm as described above.

Furthermore, the invention provides nucleic acid which can hybridise to the *S.agalactiae* nucleic acid disclosed in the examples, and nucleic acid which can hybridise to the *S.pyogenes* nucleic acid disclosed in the examples preferably under 'high stringency' conditions (*e.g.* 65°C in 0.1xSSC, 0.5% SDS solution).

Nucleic acid comprising fragments of these sequences are also provided. These should comprise at least *n* consecutive nucleotides from the *S.agalactiae* or *S.pyogenes* sequences and, depending on the particular sequence, *n* is 10 or more (*e.g.* 12, 14, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). The fragments may comprise sequences which are common to the related GAS and GBS sequences disclosed in the examples.

According to a further aspect, the invention provides nucleic acid encoding the proteins and protein fragments of the invention.

The invention also provides: nucleic acid comprising nucleotide sequence SEQ ID 10967; nucleic acid comprising nucleotide sequences having sequence identity to SEQ ID 10967; nucleic acid which can hybridise to SEQ ID 10967 (preferably under 'high stringency' conditions); nucleic acid comprising a fragment of at least *n* consecutive nucleotides from SEQ ID 10967, wherein *n* is 10 or more *e.g.* 12, 14, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1500, 2000, 3000, 4000, 5000, 10000, 100000, 1000000 or more

Nucleic acids of the invention can be used in hybridisation reactions (*e.g.* Northern or Southern blots, or in nucleic acid microarrays or 'gene chips') and amplification reactions (*e.g.* PCR, SDA, SSSR, LCR, TMA, NASBA *etc.*) and other nucleic acid techniques.

It should also be appreciated that the invention provides nucleic acid comprising sequences complementary to those described above (*e.g.* for antisense or probing, or for use as primers).

Nucleic acid according to the invention can, of course, be prepared in many ways (*e.g.* by chemical synthesis, from genomic or cDNA libraries, from the organism itself *etc.*) and can take various forms (*e.g.* single stranded, double stranded, vectors, primers, probes, labelled *etc.*). The nucleic acid is preferably in substantially isolated form.

Nucleic acid according to the invention may be labelled *e.g.* with a radioactive or fluorescent label. This is particularly useful where the nucleic acid is to be used in nucleic acid detection techniques *e.g.* where the nucleic acid is a primer or as a probe for use in techniques such as PCR, LCR, TMA, NASBA *etc.*

In addition, the term "nucleic acid" includes DNA and RNA, and also their analogues, such as those containing modified backbones, and also peptide nucleic acids (PNA) *etc.*

According to a further aspect, the invention provides vectors comprising nucleotide sequences of the invention (*e.g.* cloning or expression vectors) and host cells transformed with such vectors.

According to a further aspect, the invention provides compositions comprising protein, antibody, and/or nucleic acid according to the invention. These compositions may be suitable as immunogenic compositions, for instance, or as diagnostic reagents, or as vaccines.

The invention also provides nucleic acid, protein, or antibody according to the invention for use as medicaments (*e.g.* as immunogenic compositions or as vaccines) or as diagnostic reagents. It also provides the use of nucleic acid, protein, or antibody according to the invention in the manufacture of: (i) a medicament for treating or preventing disease and/or infection caused by streptococcus; (ii) a diagnostic reagent for detecting the presence of streptococcus or of antibodies raised against streptococcus; and/or (iii) a reagent which can raise antibodies against streptococcus. Said streptococcus may be any species, group or strain, but is preferably *S.agalactiae*, especially serotype III or V, or *S.pyogenes*. Said disease may be bacteremia, meningitis, puerperal fever, scarlet fever, erysipelas, pharyngitis, impetigo, necrotising fasciitis, myositis or toxic shock syndrome.

The invention also provides a method of treating a patient, comprising administering to the patient a therapeutically effective amount of nucleic acid, protein, and/or antibody of the invention. The patient may either be at risk from the disease themselves or may be a pregnant woman ('maternal immunisation' *e.g.* Glezen & Alpers (1999) *Clin. Infect. Dis.* 28:219-224).

Administration of protein antigens is a preferred method of treatment for inducing immunity.

Administration of antibodies of the invention is another preferred method of treatment. This method of passive immunisation is particularly useful for newborn children or for pregnant women. This method will typically use monoclonal antibodies, which will be humanised or fully human.

The invention also provides a kit comprising primers (*e.g.* PCR primers) for amplifying a template sequence contained within a *Streptococcus* (*e.g. S.pyogenes* or *S.agalactiae*) nucleic acid sequence, the kit comprising a first primer and a second primer, wherein the first primer is substantially complementary to said template sequence and the second primer is substantially complementary to a complement of said template sequence, wherein the parts of said primers which have substantial complementarity define the termini of the template sequence to be amplified. The first primer and/or the second primer may include a detectable label (*e.g.* a fluorescent label).

The invention also provides a kit comprising first and second single-stranded oligonucleotides which allow amplification of a *Streptococcus* template nucleic acid sequence contained in a single- or double-stranded nucleic acid (or mixture thereof), wherein: (a) the first oligonucleotide comprises a primer sequence which is substantially complementary to said template nucleic acid sequence; (b) the second oligonucleotide comprises a primer sequence which is substantially complementary to the complement of said template nucleic acid sequence; (c) the first oligonucleotide and/or the second oligonucleotide comprise(s) sequence which is not compementary to said template nucleic acid; and (d) said primer sequences define the termini of the template sequence to be amplified. The non-complementary sequence(s) of feature (c) are preferably upstream of (*i.e.* 5' to) the primer sequences. One or both of these (c) sequences may comprise a restriction site (*e.g.* EP-B-0509612) or a promoter sequence (*e.g.* EP-B-0505012). The first oligonucleotide and/or the second oligonucleotide may include a detectable label (*e.g.* a fluorescent label).

The template sequence may be any part of a genome sequence (*e.g.* SEQ ID 10967). For example, it could be a rRNA gene (*e.g.* Turenne *et al.* (2000) *J. Clin. Microbiol.* 38:513-520; SEQ IDs 12018-12024 herein) or a protein-coding gene. The template sequence is preferably specific to GBS.

The invention also provides a computer-readable medium (*e.g.* a floppy disk, a hard disk, a CD-ROM, a DVD *etc.*) and/or a computer database containing one or more of the sequences in the sequence listing. The medium preferably contains SEQ ID 10967.

The invention also provides a hybrid protein represented by the formula NH_2-A-[-X-L-]$_n$-B-COOH, wherein X is a protein of the invention, L is an optional linker amino acid sequence, A is an optional N-terminal amino acid sequence, B is an optional C-terminal amino acid sequence, and n is an integer greater than 1. The value of n is between 2 and x, and the value of x is typically 3, 4, 5, 6, 7, 8, 9 or 10. Preferably n is 2, 3 or 4; it is more preferably 2 or 3; most preferably, n = 2. For each n instances, -X- may be the same or different. For each n instances of [-X-L-], linker amino acid sequence -L- may be present or absent. For instance, when n=2 the hybrid may be NH_2-X_1-L_1-X_2-L_2-COOH, NH_2-X_1-X_2-COOH, NH_2-X_1-L_1-X_2-COOH, NH_2-X_1-X_2-L_2-COOH, *etc.* Linker amino acid sequence(s) -L- will typically be short (*e.g.* 20 or fewer amino acids *i.e.* 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1). Examples include short peptide sequences which facilitate cloning, poly-glycine linkers (*i.e.* Gly$_n$ where n = 2, 3, 4, 5, 6, 7, 8, 9, 10 or more), and histidine tags (*i.e.* His$_n$ where n = 3, 4, 5, 6, 7, 8, 9, 10 or more). Other suitable linker amino acid sequences will be apparent to those skilled in the art. -A- and -B- are optional sequences which will typically be short (*e.g.* 40 or fewer amino acids *i.e.* 39, 38, 37, 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1). Examples include leader sequences to direct protein trafficking, or short peptide sequences which facilitate cloning or purification (*e.g.* histidine tags *i.e.* His$_n$ where n = 3, 4, 5, 6, 7, 8, 9, 10 or more). Other suitable N-terminal and C-terminal amino acid sequences will be apparent to those

skilled in the art. In some embodiments, each X will be a GBS sequence; in others, mixtures of GAS and GBS will be used.

According to further aspects, the invention provides various processes.

A process for producing proteins of the invention is provided, comprising the step of culturing a host cell of to the invention under conditions which induce protein expression.

A process for producing protein or nucleic acid of the invention is provided, wherein the protein or nucleic acid is synthesised in part or in whole using chemical means.

A process for detecting polynucleotides of the invention is provided, comprising the steps of: (a) contacting a nucleic probe according to the invention with a biological sample under hybridising conditions to form duplexes; and (b) detecting said duplexes.

A process for detecting *Streptococcus* in a biological sample (*e.g.* blood) is also provided, comprising the step of contacting nucleic acid according to the invention with the biological sample under hybridising conditions. The process may involve nucleic acid amplification (*e.g.* PCR, SDA, SSSR, LCR, TMA, NASBA *etc.*) or hybridisation (*e.g.* microarrays, blots, hybridisation with a probe in solution *etc.*). PCR detection of *Streptococcus* in clinical samples, in particular *S.pyogenes*, has been reported [see *e.g.* Louie *et al.* (2000) *CMAJ* 163:301-309; Louie *et al.* (1998) *J. Clin. Microbiol.* 36:1769-1771]. Clinical assays based on nucleic acid are described in general in Tang *et al.* (1997) *Clin. Chem.* 43:2021-2038.

A process for detecting proteins of the invention is provided, comprising the steps of: (a) contacting an antibody of the invention with a biological sample under conditions suitable for the formation of an antibody-antigen complexes; and (b) detecting said complexes.

A process for identifying an amino acid sequence is provided, comprising the step of searching for putative open reading frames or protein-coding regions within a genome sequence of *S.agalactiae*. This will typically involve *in silico* searching the sequence for an initiation codon and for an in-frame termination codon in the downstream sequence. The region between these initiation and termination codons is a putative protein-coding sequence. Typically, all six possible reading frames will be searched. Suitable software for such analysis includes ORFFINDER (NCBI), GENEMARK [Borodovsky & McIninch (1993) *Computers Chem.* 17:122-133), GLIMMER [Salzberg *et al.* (1998) *Nucleic Acids Res.* 26:544-548; Salzberg *et al.* (1999) *Genomics* 59:24-31; Delcher *et al.* (1999) *Nucleic Acids Res.* 27:4636-4641], or other software which uses Markov models [*e.g.* Shmatkov *et al.* (1999) *Bioinformatics* 15:874-876]. The invention also provides a protein comprising the identified amino acid sequence. These proteins can then expressed using conventional techniques.

The invention also provides a process for determining whether a test compound binds to a protein of the invention. If a test compound binds to a protein of the invention and this binding inhibits the life cycle of the GBS bacterium, then the test compound can be used as an antibiotic or as a lead compound for the

design of antibiotics. The process will typically comprise the steps of contacting a test compound with a protein of the invention, and determining whether the test compound binds to said protein. Preferred proteins of the invention for use in these processes are enzymes (*e.g.* tRNA synthetases), membrane transporters and ribosomal proteins. Suitable test compounds include proteins, polypeptides, carbohydrates, lipids, nucleic acids (*e.g.* DNA, RNA, and modified forms thereof), as well as small organic compounds (*e.g.* MW between 200 and 2000 Da). The test compounds may be provided individually, but will typically be part of a library (*e.g.* a combinatorial library). Methods for detecting a binding interaction include NMR, filter-binding assays, gel-retardation assays, displacement assays, surface plasmon resonance, reverse two-hybrid *etc.* A compound which binds to a protein of the invention can be tested for antibiotic activity by contacting the compound with GBS bacteria and then monitoring for inhibition of growth. The invention also provides a compound identified using these methods.

The invention also provides a composition comprising a protein or the invention and one or more of the following antigens:

- a protein antigen from *Helicobacter pylori* such as VacA, CagA, NAP, HopX, HopY [*e.g.* WO98/04702] and/or urease.
- a protein antigen from *N.meningitidis* serogroup B, such as those in WO99/24578, WO99/36544, WO99/57280, WO00/22430, Tettelin *et al.* (2000) *Science* 287:1809-1815, Pizza *et al.* (2000) *Science* 287:1816-1820 and WO96/29412, with protein '287' and derivatives being particularly preferred.
- an outer-membrane vesicle (OMV) preparation from *N.meningitidis* serogroup B, such as those disclosed in WO01/52885; Bjune *et al.* (1991) *Lancet* 338(8775):1093-1096; Fukasawa *et al.* (1999) *Vaccine* 17:2951-2958; Rosenqvist *et al.* (1998) *Dev. Biol. Stand.* 92:323-333 *etc.*
- a saccharide antigen from *N.meningitidis* serogroup A, C, W135 and/or Y, such as the oligosaccharide disclosed in Costantino *et al.* (1992) *Vaccine* 10:691-698from serogroup C [see also Costantino *et al.* (1999) *Vaccine* 17:1251-1263].
- a saccharide antigen from *Streptococcus pneumoniae* [*e.g.* Watson (2000) *Pediatr Infect Dis J* 19:331-332; Rubin (2000) *Pediatr Clin North Am* 47:269-285, v; Jedrzejas (2001) *Microbiol Mol Biol Rev* 65:187-207].
- an antigen from hepatitis A virus, such as inactivated virus [*e.g.* Bell (2000) *Pediatr Infect Dis J* 19:1187-1188; Iwarson (1995) *APMIS* 103:321-326].
- an antigen from hepatitis B virus, such as the surface and/or core antigens [*e.g.* Gerlich *et al.* (1990) *Vaccine* 8 Suppl:S63-68 & 79-80].
- an antigen from hepatitis C virus [*e.g.* Hsu *et al.* (1999) *Clin Liver Dis* 3:901-915].
- an antigen from *Bordetella pertussis*, such as pertussis holotoxin (PT) and filamentous haemagglutinin (FHA) from *B.pertussis*, optionally also in combination with pertactin and/or

agglutinogens 2 and 3 [*e.g.* Gustafsson *et al.* (1996) *N. Engl. J. Med.* 334:349-355; Rappuoli *et al.* (1991) *TIBTECH* 9:232-238].

- a diphtheria antigen, such as a diphtheria toxoid [*e.g.* chapter 3 of *Vaccines* (1988) eds. Plotkin & Mortimer. ISBN 0-7216-1946-0] *e.g.* the CRM_{197} mutant [*e.g.* Del Guidice *et al.* (1998) *Molecular Aspects of Medicine* 19:1-70].
- a tetanus antigen, such as a tetanus toxoid [*e.g.* chapter 4 of Plotkin & Mortimer].
- a saccharide antigen from *Haemophilus influenzae* B.
- an antigen from *N.gonorrhoeae* [*e.g.* WO99/24578, WO99/36544, WO99/57280].
- an antigen from *Chlamydia pneumoniae* [*e.g.* PCT/IB01/01445; Kalman *et al.* (1999) *Nature Genetics* 21:385-389; Read *et al.* (2000) *Nucleic Acids Res* 28:1397-406; Shirai *et al.* (2000) *J. Infect. Dis.* 181(Suppl 3):S524-S527; WO99/27105; WO00/27994; WO00/37494].
- an antigen from *Chlamydia trachomatis* [*e.g.* WO99/28475].
- an antigen from *Porphyromonas gingivalis* [*e.g.* Ross *et al.* (2001) *Vaccine* 19:4135-4142].
- polio antigen(s) [*e.g.* Sutter *et al.* (2000) *Pediatr Clin North Am* 47:287-308; Zimmerman & Spann (1999) *Am Fam Physician* 59:113-118, 125-126] such as IPV or OPV.
- rabies antigen(s) [*e.g.* Dreesen (1997) *Vaccine* 15 Suppl:S2-6] such as lyophilised inactivated virus [*e.g. MMWR Morb Mortal Wkly Rep* 1998 Jan 16;47(1):12, 19; RabAvert™].
- measles, mumps and/or rubella antigens [*e.g.* chapters 9, 10 & 11 of Plotkin & Mortimer].
- influenza antigen(s) [*e.g.* chapter 19 of Plotkin & Mortimer], such as the haemagglutinin and/or neuraminidase surface proteins.
- an antigen from *Moraxella catarrhalis* [*e.g.* McMichael (2000) *Vaccine* 19 Suppl 1:S101-107].
- an antigen from *Staphylococcus aureus* [*e.g.* Kuroda *et al.* (2001) *Lancet* 357(9264):1225-1240; see also pages 1218-1219].

Where a saccharide or carbohydrate antigen is included, it is preferably conjugated to a carrier protein in order to enhance immunogenicity [*e.g.* Ramsay *et al.* (2001) *Lancet* 357(9251):195-196; Lindberg (1999) *Vaccine* 17 Suppl 2:S28-36; *Conjugate Vaccines* (eds. Cruse *et al.*) ISBN 3805549326, particularly vol. 10:48-114 *etc.*]. Preferred carrier proteins are bacterial toxins or toxoids, such as diphtheria or tetanus toxoids. The CRM_{197} diphtheria toxoid is particularly preferred. Other suitable carrier proteins include the *N.meningitidis* outer membrane protein [*e.g.* EP-0372501], synthetic peptides [*e.g.* EP-0378881, EP-0427347], heat shock proteins [*e.g.* WO93/17712], pertussis proteins [*e.g.* WO98/58668; EP-0471177], protein D from *H.influenzae* [*e.g.* WO00/56360], toxin A or B from *C.difficile* [*e.g.* WO00/61761], *etc.* Any suitable conjugation reaction can be used, with any suitable linker where necessary.

Toxic protein antigens may be detoxified where necessary (*e.g.* detoxification of pertussis toxin by chemical and/or genetic means).

Where a diphtheria antigen is included in the composition it is preferred also to include tetanus antigen and pertussis antigens. Similarly, where a tetanus antigen is included it is preferred also to include diphtheria and pertussis antigens. Similarly, where a pertussis antigen is included it is preferred also to include diphtheria and tetanus antigens.

Antigens are preferably adsorbed to an aluminium salt.

Antigens in the composition will typically be present at a concentration of at least 1μg/ml each. In general, the concentration of any given antigen will be sufficient to elicit an immune response against that antigen.

The invention also provides compositions comprising two or more proteins of the present invention. The two or more proteins may comprise GBS sequences or may comprise GAS and GBS sequences.

A summary of standard techniques and procedures which may be employed to perform the invention (*e.g.* to utilise the disclosed sequences for vaccination or diagnostic purposes) follows. This summary is not a limitation on the invention but, rather, gives examples that may be used, but are not required.

General

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature *eg.* Sambrook *Molecular Cloning; A Laboratory Manual, Second Edition* (1989); *DNA Cloning, Volumes I and II* (D.N Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed, 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds. 1984); *Transcription and Translation* (B.D. Hames & S.J. Higgins eds. 1984); *Animal Cell Culture* (R.I. Freshney ed. 1986); *Immobilized Cells and Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide to Molecular Cloning* (1984); the *Methods in Enzymology* series (Academic Press, Inc.), especially volumes 154 & 155; *Gene Transfer Vectors for Mammalian Cells* (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Mayer and Walker, eds. (1987), *Immunochemical Methods in Cell and Molecular Biology* (Academic Press, London); Scopes, (1987) *Protein Purification: Principles and Practice*, Second Edition (Springer-Verlag, N.Y.), and *Handbook of Experimental Immunology, Volumes I-IV* (D.M. Weir and C. C. Blackwell eds 1986).

Standard abbreviations for nucleotides and amino acids are used in this specification.

Definitions

A composition containing X is "substantially free of" Y when at least 85% by weight of the total X+Y in the composition is X. Preferably, X comprises at least about 90% by weight of the total of X+Y in the composition, more preferably at least about 95% or even 99% by weight.

The term "comprising" means "including" as well as "consisting" *e.g.* a composition "comprising" X may consist exclusively of X or may include something additional *e.g.* X + Y.

The term "heterologous" refers to two biological components that are not found together in nature. The components may be host cells, genes, or regulatory regions, such as promoters. Although the heterologous components are not found together in nature, they can function together, as when a promoter heterologous to a gene is operably linked to the gene. Another example is where a streptococcus sequence is heterologous to a mouse host cell. A further examples would be two epitopes from the same or different proteins which have been assembled in a single protein in an arrangement not found in nature

An "origin of replication" is a polynucleotide sequence that initiates and regulates replication of polynucleotides, such as an expression vector. The origin of replication behaves as an autonomous unit of polynucleotide replication within a cell, capable of replication under its own control. An origin of replication may be needed for a vector to replicate in a particular host cell. With certain origins of replication, an expression vector can be reproduced at a high copy number in the presence of the appropriate proteins within the cell. Examples of origins are the autonomously replicating sequences, which are effective in yeast; and the viral T-antigen, effective in COS-7 cells.

A "mutant" sequence is defined as DNA, RNA or amino acid sequence differing from but having sequence identity with the native or disclosed sequence. Depending on the particular sequence, the degree of sequence identity between the native or disclosed sequence and the mutant sequence is preferably greater than 50% (*eg.* 60%, 70%, 80%, 90%, 95%, 99% or more, calculated using the Smith-Waterman algorithm as described above). As used herein, an "allelic variant" of a nucleic acid molecule, or region, for which nucleic acid sequence is provided herein is a nucleic acid molecule, or region, that occurs essentially at the same locus in the genome of another or second isolate, and that, due to natural variation caused by, for example, mutation or recombination, has a similar but not identical nucleic acid sequence. A coding region allelic variant typically encodes a protein having similar activity to that of the protein encoded by the gene to which it is being compared. An allelic variant can also comprise an alteration in the 5' or 3' untranslated regions of the gene, such as in regulatory control regions (*eg.* see US patent 5,753,235).

Expression systems

The streptococcus nucleotide sequences can be expressed in a variety of different expression systems; for example those used with mammalian cells, baculoviruses, plants, bacteria, and yeast.

i. Mammalian Systems

Mammalian expression systems are known in the art. A mammalian promoter is any DNA sequence capable of binding mammalian RNA polymerase and initiating the downstream (3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiating region, which is usually placed proximal to the 5' end of the coding sequence, and a TATA box, usually located 25-30 base pairs (bp) upstream of the transcription initiation site. The TATA box is thought to direct RNA polymerase II to begin RNA synthesis at the correct site. A mammalian promoter will also contain an upstream promoter element, usually located within 100 to 200 bp upstream of the TATA box. An upstream promoter element determines the rate at which transcription is initiated and can act in either orientation [Sambrook et al. (1989) "Expression of Cloned Genes in Mammalian Cells." In *Molecular Cloning: A Laboratory Manual, 2nd ed.]*.

Mammalian viral genes are often highly expressed and have a broad host range; therefore sequences encoding mammalian viral genes provide particularly useful promoter sequences. Examples include the SV40 early promoter, mouse mammary tumor virus LTR promoter, adenovirus major late promoter (Ad MLP), and herpes simplex virus promoter. In addition, sequences derived from non-viral genes, such as the murine metallotheionein gene, also provide useful promoter sequences. Expression may be either constitutive or regulated (inducible), depending on the promoter can be induced with glucocorticoid in hormone-responsive cells.

The presence of an enhancer element (enhancer), combined with the promoter elements described above, will usually increase expression levels. An enhancer is a regulatory DNA sequence that can stimulate transcription up to 1000-fold when linked to homologous or heterologous promoters, with synthesis beginning at the normal RNA start site. Enhancers are also active when they are placed upstream or downstream from the transcription initiation site, in either normal or flipped orientation, or at a distance of more than 1000 nucleotides from the promoter [Maniatis et al. (1987) *Science 236:*1237; Alberts et al. (1989) *Molecular Biology of the Cell*, 2nd ed.]. Enhancer elements derived from viruses may be particularly useful, because they usually have a broader host range. Examples include the SV40 early gene enhancer [Dijkema et al (1985) *EMBO J. 4:*761] and the enhancer/promoters derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus [Gorman et al. (1982b) *Proc. Natl. Acad. Sci. 79:*6777] and from human cytomegalovirus [Boshart et al. (1985) *Cell 41:*521]. Additionally, some enhancers are regulatable and become active only in the presence of an inducer, such as a hormone or metal ion [Sassone-Corsi and Borelli (1986) *Trends Genet. 2:*215; Maniatis et al. (1987) *Science* 236:1237].

A DNA molecule may be expressed intracellularly in mammalian cells. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in mammalian

cells. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The adenovirus triparite leader is an example of a leader sequence that provides for secretion of a foreign protein in mammalian cells.

Usually, transcription termination and polyadenylation sequences recognized by mammalian cells are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-transcriptional cleavage and polyadenylation [Birnstiel et al. (1985) *Cell 41*:349; Proudfoot and Whitelaw (1988) "Termination and 3' end processing of eukaryotic RNA. In *Transcription and splicing* (ed. B.D. Hames and D.M. Glover); Proudfoot (1989) *Trends Biochem. Sci. 14*:105]. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminater/polyadenylation signals include those derived from SV40 [Sambrook et al (1989) "Expression of cloned genes in cultured mammalian cells." In *Molecular Cloning: A Laboratory Manual*].

Usually, the above described components, comprising a promoter, polyadenylation signal, and transcription termination sequence are put together into expression constructs. Enhancers, introns with functional splice donor and acceptor sites, and leader sequences may also be included in an expression construct, if desired. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (*eg.* plasmids) capable of stable maintenance in a host, such as mammalian cells or bacteria. Mammalian replication systems include those derived from animal viruses, which require trans-acting factors to replicate. For example, plasmids containing the replication systems of papovaviruses, such as SV40 [Gluzman (1981) *Cell 23*:175] or polyomavirus, replicate to extremely high copy number in the presence of the appropriate viral T antigen. Additional examples of mammalian replicons include those derived from bovine papillomavirus and Epstein-Barr virus. Additionally, the replicon may have two replicaton systems, thus allowing it to be maintained, for example, in mammalian cells for expression and in a prokaryotic host for cloning and amplification. Examples of such mammalian-bacteria shuttle vectors include pMT2 [Kaufman et al. (1989) *Mol. Cell. Biol. 9*:946] and pHEBO [Shimizu et al. (1986) *Mol. Cell. Biol. 6*:1074].

The transformation procedure used depends upon the host to be transformed. Methods for introduction of heterologous polynucleotides into mammalian cells are known in the art and include dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (*eg.* Hep G2), and a number of other cell lines.

ii. Baculovirus Systems

The polynucleotide encoding the protein can also be inserted into a suitable insect expression vector, and is operably linked to the control elements within that vector. Vector construction employs techniques which are known in the art. Generally, the components of the expression system include a transfer vector, usually a bacterial plasmid, which contains both a fragment of the baculovirus genome, and a convenient restriction site for insertion of the heterologous gene or genes to be expressed; a wild type baculovirus with a sequence homologous to the baculovirus-specific fragment in the transfer vector (this allows for the homologous recombination of the heterologous gene in to the baculovirus genome); and appropriate insect host cells and growth media.

After inserting the DNA sequence encoding the protein into the transfer vector, the vector and the wild type viral genome are transfected into an insect host cell where the vector and viral genome are allowed to recombine. The packaged recombinant virus is expressed and recombinant plaques are identified and purified. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA ("MaxBac" kit). These techniques are generally known to those skilled in the art and fully described in Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987) (hereinafter "Summers and Smith").

Prior to inserting the DNA sequence encoding the protein into the baculovirus genome, the above described components, comprising a promoter, leader (if desired), coding sequence, and transcription termination sequence, are usually assembled into an intermediate transplacement construct (transfer vector). This may contain a single gene and operably linked regulatory elements; multiple genes, each with its owned set of operably linked regulatory elements; or multiple genes, regulated by the same set of regulatory elements. Intermediate transplacement constructs are often maintained in a replicon, such as an extra-chromosomal

element (*e.g.* plasmids) capable of stable maintenance in a host, such as a bacterium. The replicon will have a replication system, thus allowing it to be maintained in a suitable host for cloning and amplification.

Currently, the most commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed. These include, for example, pVL985 (which alters the polyhedrin start codon from ATG to ATT, and which introduces a BamHI cloning site 32 basepairs downstream from the ATT; see Luckow and Summers, *Virology* (1989) *17*:31.

The plasmid usually also contains the polyhedrin polyadenylation signal (Miller et al. (1988) *Ann. Rev. Microbiol.*, *42*:177) and a prokaryotic ampicillin-resistance (*amp*) gene and origin of replication for selection and propagation in *E.coli*.

Baculovirus transfer vectors usually contain a baculovirus promoter. A baculovirus promoter is any DNA sequence capable of binding a baculovirus RNA polymerase and initiating the downstream (5' to 3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A baculovirus transfer vector may also have a second domain called an enhancer, which, if present, is usually distal to the structural gene. Expression may be either regulated or constitutive.

Structural genes, abundantly transcribed at late times in a viral infection cycle, provide particularly useful promoter sequences. Examples include sequences derived from the gene encoding the viral polyhedron protein, Friesen et al., (1986) "The Regulation of Baculovirus Gene Expression," in: *The Molecular Biology of Baculoviruses* (ed. Walter Doerfler); EPO Publ. Nos. 127 839 and 155 476; and the gene encoding the p10 protein, Vlak et al., (1988), *J. Gen. Virol. 69*:765.

DNA encoding suitable signal sequences can be derived from genes for secreted insect or baculovirus proteins, such as the baculovirus polyhedrin gene (Carbonell et al. (1988) *Gene*, *73*:409). Alternatively, since the signals for mammalian cell posttranslational modifications (such as signal peptide cleavage, proteolytic cleavage, and phosphorylation) appear to be recognized by insect cells, and the signals required for secretion and nuclear accumulation also appear to be conserved between the invertebrate cells and vertebrate cells, leaders of non-insect origin, such as those derived from genes encoding human α-interferon, Maeda et al., (1985), *Nature 315*:592; human gastrin-releasing peptide, Lebacq-Verheyden et al., (1988), *Molec. Cell. Biol. 8*:3129; human IL-2, Smith et al., (1985) *Proc. Nat'l Acad. Sci. USA*, *82*:8404; mouse IL-3, (Miyajima et al., (1987) *Gene 58*:273; and human glucocerebrosidase, Martin et al. (1988) *DNA*, *7*:99, can also be used to provide for secretion in insects.

A recombinant polypeptide or polyprotein may be expressed intracellularly or, if it is expressed with the proper regulatory sequences, it can be secreted. Good intracellular expression of nonfused foreign proteins usually requires heterologous genes that ideally have a short leader sequence containing suitable translation initiation signals preceding an ATG start signal. If desired, methionine at the N-terminus may be cleaved from the mature protein by *in vitro* incubation with cyanogen bromide.

Alternatively, recombinant polyproteins or proteins which are not naturally secreted can be secreted from the insect cell by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in insects. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the translocation of the protein into the endoplasmic reticulum.

After insertion of the DNA sequence and/or the gene encoding the expression product precursor of the protein, an insect cell host is co-transformed with the heterologous DNA of the transfer vector and the genomic DNA of wild type baculovirus -- usually by co-transfection. The promoter and transcription termination sequence of the construct will usually comprise a 2-5kb section of the baculovirus genome. Methods for introducing heterologous DNA into the desired site in the baculovirus virus are known in the art. (See Summers and Smith *supra*; Ju et al. (1987); Smith et al., *Mol. Cell. Biol.* (1983) *3*:2156; and Luckow and Summers (1989)). For example, the insertion can be into a gene such as the polyhedrin gene, by homologous double crossover recombination; insertion can also be into a restriction enzyme site engineered into the desired baculovirus gene. Miller et al., (1989), *Bioessays 4*:91.The DNA sequence, when cloned in place of the polyhedrin gene in the expression vector, is flanked both 5' and 3' by polyhedrin-specific sequences and is positioned downstream of the polyhedrin promoter.

The newly formed baculovirus expression vector is subsequently packaged into an infectious recombinant baculovirus. Homologous recombination occurs at low frequency (between about 1% and about 5%); thus, the majority of the virus produced after cotransfection is still wild-type virus. Therefore, a method is necessary to identify recombinant viruses. An advantage of the expression system is a visual screen allowing recombinant viruses to be distinguished. The polyhedrin protein, which is produced by the native virus, is produced at very high levels in the nuclei of infected cells at late times after viral infection. Accumulated polyhedrin protein forms occlusion bodies that also contain embedded particles. These occlusion bodies, up to 15 μm in size, are

highly refractile, giving them a bright shiny appearance that is readily visualized under the light microscope. Cells infected with recombinant viruses lack occlusion bodies. To distinguish recombinant virus from wild-type virus, the transfection supernatant is plaqued onto a monolayer of insect cells by techniques known to those skilled in the art. Namely, the plaques are screened under the light microscope for the presence (indicative of wild-type virus) or absence (indicative of recombinant virus) of occlusion bodies. "Current Protocols in Microbiology" Vol. 2 (Ausubel et al. eds) at 16.8 (Supp. 10, 1990); Summers and Smith, *supra*; Miller et al. (1989).

Recombinant baculovirus expression vectors have been developed for infection into several insect cells. For example, recombinant baculoviruses have been developed for, *inter alia*: *Aedes aegypti* , *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni* (WO 89/046699; Carbonell et al., (1985) *J. Virol. 56*:153; Wright (1986) *Nature 321*:718; Smith et al., (1983) *Mol. Cell. Biol. 3*:2156; and see generally, Fraser, *et al.* (1989) *In Vitro Cell. Dev. Biol. 25*:225).

Cells and cell culture media are commercially available for both direct and fusion expression of heterologous polypeptides in a baculovirus/expression system; cell culture technology is generally known to those skilled in the art. *See, eg.* Summers and Smith *supra*.

The modified insect cells may then be grown in an appropriate nutrient medium, which allows for stable maintenance of the plasmid(s) present in the modified insect host. Where the expression product gene is under inducible control, the host may be grown to high density, and expression induced. Alternatively, where expression is constitutive, the product will be continuously expressed into the medium and the nutrient medium must be continuously circulated, while removing the product of interest and augmenting depleted nutrients. The product may be purified by such techniques as chromatography, *eg.* HPLC, affinity chromatography, ion exchange chromatography, *etc.*; electrophoresis; density gradient centrifugation; solvent extraction, *etc.* As appropriate, the product may be further purified, as required, so as to remove substantially any insect proteins which are also present in the medium, so as to provide a product which is at least substantially free of host debris, eg. proteins, lipids and polysaccharides.

In order to obtain protein expression, recombinant host cells derived from the transformants are incubated under conditions which allow expression of the recombinant protein encoding sequence. These conditions will vary, dependent upon the host cell selected. However, the conditions are readily ascertainable to those of ordinary skill in the art, based upon what is known in the art.

iii. Plant Systems

There are many plant cell culture and whole plant genetic expression systems known in the art. Exemplary plant cellular genetic expression systems include those described in patents, such as: US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, *Phytochemistry* 30:3861-3863 (1991). Descriptions of plant protein signal peptides may be found in addition to the references described above in Vaulcombe et al.,*Mol. Gen. Genet*. 209:33-40 (1987); Chandler et al., *Plant Molecular Biology* 3:407-418 (1984); Rogers, *J. Biol. Chem.* 260:3731-3738 (1985); Rothstein et al., *Gene* 55:353-356 (1987); Whittier et al., Nucleic Acids Research 15:2515-2535 (1987); Wirsel et al., *Molecular Microbiology* 3:3-14 (1989); Yu et al., *Gene* 122:247-253 (1992). A description of the regulation of plant gene expression by the phytohormone, gibberellic acid and secreted enzymes induced by gibberellic acid can be found in R.L. Jones and J. MacMillin, Gibberellins: in: *Advanced Plant Physiology*,. Malcolm B. Wilkins, ed., 1984 Pitman Publishing Limited, London, pp. 21-52. References that describe other metabolically-regulated genes: Sheen, *Plant Cell*, 2:1027-1038(1990); Maas et al., *EMBO J.* 9:3447-3452 (1990); Benkel and Hickey, *Proc. Natl. Acad. Sci.* 84:1337-1339 (1987).

Typically, using techniques known in the art, a desired polynucleotide sequence is inserted into an expression cassette comprising genetic regulatory elements designed for operation in plants. The expression cassette is inserted into a desired expression vector with companion sequences upstream and downstream from the expression cassette suitable for expression in a plant host. The companion sequences will be of plasmid or viral origin and provide necessary characteristics to the vector to permit the vectors to move DNA from an original cloning host, such as bacteria, to the desired plant host. The basic bacterial/plant vector construct will preferably provide a broad host range prokaryote replication origin; a prokaryote selectable marker; and, for Agrobacterium transformations, T DNA sequences for Agrobacterium-mediated transfer to plant chromosomes. Where the heterologous gene is not readily amenable to detection, the construct will preferably also have a selectable marker gene suitable for determining if a plant cell has been transformed. A general review of suitable markers, for example for the members of the grass family, is found in Wilmink and Dons, 1993, *Plant Mol. Biol. Reptr*, 11(2):165-185.

Sequences suitable for permitting integration of the heterologous sequence into the plant genome are also recommended. These might include transposon sequences and the like for homologous recombination as well as Ti sequences which permit random insertion of a heterologous expression cassette into a plant genome. Suitable prokaryote selectable markers include resistance toward antibiotics such as ampicillin or tetracycline. Other DNA sequences encoding additional functions may also be present in the vector, as is known in the art.

The nucleic acid molecules of the subject invention may be included into an expression cassette for expression of the protein(s) of interest. Usually, there will be only one expression cassette, although two or more are feasible. The recombinant expression cassette will contain in addition to the heterologous protein encoding sequence the following elements, a promoter region, plant 5' untranslated sequences, initiation codon depending upon whether or not the structural gene comes equipped with one, and a transcription and translation termination sequence. Unique restriction enzyme sites at the 5' and 3' ends of the cassette allow for easy insertion into a pre-existing vector.

A heterologous coding sequence may be for any protein relating to the present invention. The sequence encoding the protein of interest will encode a signal peptide which allows processing and translocation of the protein, as appropriate, and will usually lack any sequence which might result in the binding of the desired protein of the invention to a membrane. Since, for the most part, the transcriptional initiation region will be for a gene which is expressed and translocated during germination, by employing the signal peptide which provides for translocation, one may also provide for translocation of the protein of interest. In this way, the protein(s) of interest will be translocated from the cells in which they are expressed and may be efficiently harvested. Typically secretion in seeds are across the aleurone or scutellar epithelium layer into the endosperm of the seed. While it is not required that the protein be secreted from the cells in which the protein is produced, this facilitates the isolation and purification of the recombinant protein.

Since the ultimate expression of the desired gene product will be in a eucaryotic cell it is desirable to determine whether any portion of the cloned gene contains sequences which will be processed out as introns by the host's splicosome machinery. If so, site-directed mutagenesis of the "intron" region may be conducted to prevent losing a portion of the genetic message as a false intron code, Reed and Maniatis, *Cell* 41:95-105, 1985.

The vector can be microinjected directly into plant cells by use of micropipettes to mechanically transfer the recombinant DNA. Crossway, *Mol. Gen. Genet*, 202:179-185, 1985. The genetic material may also be transferred into the plant cell by using polyethylene glycol, Krens, et al., *Nature*, 296, 72-74, 1982. Another method of introduction of nucleic acid segments is high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface, Klein, et al., *Nature*, 327, 70-73, 1987 and Knudsen and Muller, 1991, *Planta*, 185:330-336 teaching particle bombardment of barley endosperm to create transgenic barley. Yet another method of introduction would be fusion of protoplasts with other entities, either minicells, cells, lysosomes or other fusible lipid-surfaced bodies, Fraley, et al., *Proc. Natl. Acad. Sci. USA*, 79, 1859-1863, 1982.

The vector may also be introduced into the plant cells by electroporation. (Fromm et al., *Proc. Natl Acad. Sci. USA* 82:5824, 1985). In this technique, plant protoplasts are electroporated in the presence of plasmids containing the gene construct. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form plant callus.

All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed by the present invention so that whole plants are recovered which contain the transferred gene. It is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables. Some suitable plants include, for example, species from the genera *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersion*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Cichorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Hererocallis*, *Nemesia*, *Pelargonium*, *Panicum*, *Pennisetum*, *Ranunculus*, *Senecio*, *Salpiglossis*, *Cucumis*, *Browaalia*, *Glycine*, *Lolium*, *Zea*, *Triticum*, *Sorghum*, and *Datura*.

Means for regeneration vary from species to species of plants, but generally a suspension of transformed protoplasts containing copies of the heterologous gene is first provided. Callus tissue is formed and shoots may be induced from callus and subsequently rooted. Alternatively, embryo formation can be induced from the protoplast suspension. These embryos germinate as natural embryos to form plants. The culture media will generally contain various amino acids and hormones, such as auxin and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Shoots and

roots normally develop simultaneously. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. If these three variables are controlled, then regeneration is fully reproducible and repeatable.

In some plant cell culture systems, the desired protein of the invention may be excreted or alternatively, the protein may be extracted from the whole plant. Where the desired protein of the invention is secreted into the medium, it may be collected. Alternatively, the embryos and embryoless-half seeds or other plant tissue may be mechanically disrupted to release any secreted protein between cells and tissues. The mixture may be suspended in a buffer solution to retrieve soluble proteins. Conventional protein isolation and purification methods will be then used to purify the recombinant protein. Parameters of time, temperature pH, oxygen, and volumes will be adjusted through routine methods to optimize expression and recovery of heterologous protein.

iv. Bacterial Systems

Bacterial expression techniques are known in the art. A bacterial promoter is any DNA sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A bacterial promoter may also have a second domain called an operator, that may overlap an adjacent RNA polymerase binding site at which RNA synthesis begins. The operator permits negative regulated (inducible) transcription, as a gene repressor protein may bind the operator and thereby inhibit transcription of a specific gene. Constitutive expression may occur in the absence of negative regulatory elements, such as the operator. In addition, positive regulation may be achieved by a gene activator protein binding sequence, which, if present is usually proximal (5') to the RNA polymerase binding sequence. An example of a gene activator protein is the catabolite activator protein (CAP), which helps initiate transcription of the lac operon in Escherichia coli (*E.coli*) [Raibaud *et al.* (1984) *Annu. Rev. Genet. 18*:173]. Regulated expression may therefore be either positive or negative, thereby either enhancing or reducing transcription.

Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose (*lac*) [Chang *et al.* (1977) *Nature 198*:1056], and maltose. Additional examples include promoter sequences derived from biosynthetic enzymes such as tryptophan (*trp*) [Goeddel *et al.* (1980) *Nuc. Acids Res. 8*:4057; Yelverton *et al.* (1981) *Nucl. Acids Res. 9*:731; US patent 4,738,921; EP-A-0036776 and EP-A-0121775]. The g-laotamase (*bla*) promoter system [Weissmann (1981) "The cloning of interferon and other mistakes." In *Interferon 3* (ed. I. Gresser)], bacteriophage lambda PL [Shimatake *et al.* (1981) *Nature 292*:128] and T5 [US patent 4,689,406] promoter systems also provide useful promoter sequences.

In addition, synthetic promoters which do not occur in nature also function as bacterial promoters. For example, transcription activation sequences of one bacterial or bacteriophage promoter may be joined with the operon sequences of another bacterial or bacteriophage promoter, creating a synthetic hybrid promoter [US patent 4,551,433]. For example, the *tac* promoter is a hybrid *trp-lac* promoter comprised of both *trp* promoter and *lac* operon sequences that is regulated by the *lac* repressor [Amann *et al.* (1983) *Gene 25*:167; de Boer *et al.* (1983) *Proc. Natl. Acad. Sci. 80*:21]. Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. A naturally occurring promoter of non-bacterial origin can also be coupled with a compatible RNA polymerase to produce high levels of expression of some genes in prokaryotes. The bacteriophage T7 RNA polymerase/promoter system is an example of a coupled promoter system [Studier *et al.* (1986) *J. Mol. Biol. 189*:113; Tabor *et al.* (1985) *Proc Natl. Acad. Sci. 82*:1074]. In addition, a hybrid promoter can also be comprised of a bacteriophage promoter and an *E.coli* operator region (EPO-A-0 267 851).

In addition to a functioning promoter sequence, an efficient ribosome binding site is also useful for the expression of foreign genes in prokaryotes. In *E.coli*, the ribosome binding site is called the Shine-Dalgarno (SD) sequence and includes an initiation codon (ATG) and a sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon [Shine *et al.* (1975) *Nature 254*:34]. The SD sequence is thought to promote binding of mRNA to the ribosome by the pairing of bases between the SD sequence and the 3' and of *E.coli* 16S rRNA [Steitz *et al.* (1979) "Genetic signals and nucleotide sequences in messenger RNA." In *Biological Regulation and Development: Gene Expression* (ed. R.F. Goldberger)]. To express eukaryotic genes and prokaryotic genes with weak ribosome-binding site [Sambrook *et al.* (1989) "Expression of cloned genes in Escherichia coli." In *Molecular Cloning: A Laboratory Manual*].

A DNA molecule may be expressed intracellularly. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide or by either *in vivo* on *in vitro* incubation with a bacterial methionine N-terminal peptidase (EP-A-0 219 237).

Fusion proteins provide an alternative to direct expression. Usually, a DNA sequence encoding the N-terminal portion of an endogenous bacterial protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the bacteriophage lambda cell gene can be linked at the 5' terminus of a foreign gene and expressed in bacteria. The resulting fusion protein preferably retains a site for a processing enzyme (factor Xa) to cleave the bacteriophage protein from the foreign gene [Nagai *et al.* (1984) *Nature 309*:810]. Fusion proteins can also be made with sequences from the *lacZ* [Jia *et al.* (1987) *Gene 60*:197], *trpE* [Allen et al. (1987) J. Biotechnol. 5:93; Makoff *et al.* (1989) *J. Gen. Microbiol. 135*:11], and *Chey* [EP-A-0 324 647] genes. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*eg.* ubiquitin specific processing-protease) to cleave the ubiquitin from the foreign protein. Through this method, native foreign protein can be isolated [Miller *et al.* (1989) *Bio/Technology 7*:698].

Alternatively, foreign proteins can also be secreted from the cell by creating chimeric DNA molecules that encode a fusion protein comprised of a signal peptide sequence fragment that provides for secretion of the foreign protein in bacteria [US patent 4,336,336]. The signal sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). Preferably there are processing sites, which can be cleaved either *in vivo* or *in vitro* encoded between the signal peptide fragment and the foreign gene.

DNA encoding suitable signal sequences can be derived from genes for secreted bacterial proteins, such as the *E.coli* outer membrane protein gene (*ompA*) [Masui *et al.* (1983), in: *Experimental Manipulation of Gene Expression*; Ghrayeb *et al.* (1984) *EMBO J. 3*:2437] and the *E.coli* alkaline phosphatase signal sequence (*phoA*) [Oka *et al.* (1985) *Proc. Natl. Acad. Sci. 82*:7212]. As an additional example, the signal sequence of the alpha-amylase gene from various Bacillus strains can be used to secrete heterologous proteins from *B. subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA 79*:5582; EP-A-0 244 042].

Usually, transcription termination sequences recognized by bacteria are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Transcription termination sequences frequently include DNA sequences of about 50 nucleotides capable of forming stem loop structures that aid in terminating transcription. Examples include transcription termination sequences derived from genes with strong promoters, such as the *trp* gene in *E.coli* as well as other biosynthetic genes.

Usually, the above described components, comprising a promoter, signal sequence (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (*eg.* plasmids) capable of stable maintenance in a host, such as bacteria. The replicon will have a replication system, thus allowing it to be maintained in a prokaryotic host either for expression or for cloning and amplification. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably contain at least about 10, and more preferably at least about 20 plasmids. Either a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host.

Alternatively, the expression constructs can be integrated into the bacterial genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to the bacterial chromosome that allows the vector to integrate. Integrations appear to result from recombinations between homologous DNA in the vector and the bacterial chromosome. For example, integrating vectors constructed with DNA from various Bacillus strains integrate into the Bacillus chromosome (EP-A- 0 127 328). Integrating vectors may also be comprised of bacteriophage or transposon sequences.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of bacterial strains that have been transformed. Selectable markers can be expressed in the bacterial host and may include genes which render bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin (neomycin), and tetracycline [Davies *et al.* (1978) *Annu. Rev. Microbiol. 32*:469]. Selectable markers may also include biosynthetic genes, such as those in the histidine, tryptophan, and leucine biosynthetic pathways.

Alternatively, some of the above described components can be put together in transformation vectors. Transformation vectors are usually comprised of a selectable market that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extra-chromosomal replicons or integrating vectors, have been developed for transformation into many bacteria. For example, expression vectors have been developed for, *inter alia*, the following bacteria: Bacillus subtilis [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA 79*:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541], Escherichia coli [Shimatake *et al.* (1981) *Nature 292*:128; Amann *et al.* (1985) *Gene 40*:183; Studier *et al.* (1986) *J. Mol. Biol. 189*:113; EP-A-0 036 776,EP-A-0 136 829 and EP-A-0 136 907], Streptococcus cremoris [Powell *et al.* (1988) *Appl. Environ. Microbiol. 54*:655]; Streptococcus lividans [Powell *et al.* (1988) *Appl. Environ. Microbiol. 54*:655], Streptomyces lividans [US patent 4,745,056].

Methods of introducing exogenous DNA into bacterial hosts are well-known in the art, and usually include either the transformation of bacteria treated with $CaCl_2$ or other agents, such as divalent cations and DMSO. DNA can also be introduced into bacterial cells by electroporation. Transformation procedures usually vary with the bacterial species to be transformed. See *eg.* [Masson *et al.* (1989) *FEMS Microbiol. Lett. 60*:273; Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA 79*:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541, Bacillus], [Miller *et al.* (1988) *Proc. Natl. Acad. Sci. 85*:856; Wang *et al.* (1990) *J. Bacteriol. 172*:949, Campylobacter], [Cohen *et al.* (1973) *Proc. Natl. Acad. Sci. 69*:2110; Dower *et al.* (1988) *Nucleic Acids Res. 16*:6127; Kushner (1978) "An improved method for transformation of Escherichia coli with ColE1-derived plasmids. In *Genetic Engineering: Proceedings of the International Symposium on Genetic Engineering* (eds. H.W. Boyer and S. Nicosia); Mandel *et al.* (1970) *J. Mol. Biol. 53*:159; Taketo (1988) *Biochim. Biophys. Acta 949*:318; Escherichia], [Chassy *et al.* (1987) *FEMS Microbiol. Lett. 44*:173 Lactobacillus]; [Fiedler *et al.* (1988) *Anal. Biochem 170*:38, Pseudomonas]; [Augustin *et al.* (1990) *FEMS Microbiol. Lett. 66*:203, Staphylococcus], [Barany *et al.* (1980) *J. Bacteriol. 144*:698; Harlander (1987) "Transformation of Streptococcus lactis by electroporation, in: *Streptococcal Genetics* (ed. J. Ferretti and R. Curtiss III); Perry *et al.* (1981) *Infect. Immun. 32*:1295; Powell *et al.* (1988) *Appl. Environ. Microbiol. 54*:655; Somkuti *et al.* (1987) *Proc. 4th Evr. Cong. Biotechnology 1*:412, Streptococcus].

v. Yeast Expression

Yeast expression systems are also known to one of ordinary skill in the art. A yeast promoter is any DNA sequence capable of binding yeast RNA polymerase and initiating the downstream (3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site (the "TATA Box") and a transcription initiation site. A yeast promoter may also have a second domain called an upstream activator sequence (UAS), which, if present, is usually distal to the structural gene. The UAS permits regulated (inducible) expression. Constitutive expression occurs in the absence of a UAS. Regulated expression may be either positive or negative, thereby either enhancing or reducing transcription.

Yeast is a fermenting organism with an active metabolic pathway, therefore sequences encoding enzymes in the metabolic pathway provide particularly useful promoter sequences. Examples include alcohol dehydrogenase (ADH) (EP-A-0 284 044), enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH), hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, and pyruvate kinase (PyK) (EPO-A-0 329 203). The yeast *PHO5* gene, encoding acid phosphatase, also provides useful promoter sequences [Myanohara *et al.* (1983) *Proc. Natl. Acad. Sci. USA 80*:1].

In addition, synthetic promoters which do not occur in nature also function as yeast promoters. For example, UAS sequences of one yeast promoter may be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region (US Patent Nos. 4,876,197 and 4,880,734). Other examples of hybrid promoters include promoters which consist of the regulatory sequences of either the *ADH2*, *GAL4*, *GAL10*, OR *PHO5* genes, combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK (EP-A-0 164 556). Furthermore, a yeast promoter can include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription. Examples of such promoters include, *inter alia*, [Cohen *et al.* (1980) *Proc. Natl. Acad. Sci. USA 77*:1078; Henikoff *et al.* (1981) *Nature 283*:835; Hollenberg *et al.* (1981) *Curr. Topics Microbiol. Immunol. 96*:119; Hollenberg *et al.* (1979) "The Expression of Bacterial Antibiotic Resistance Genes in the Yeast Saccharomyces cerevisiae," in: *Plasmids of Medical, Environmental and Commercial Importance* (eds. K.N. Timmis and A. Puhler); Mercerau-Puigalon *et al.* (1980) *Gene 11*:163; Panthier *et al.* (1980) *Curr. Genet. 2*:109;].

A DNA molecule may be expressed intracellularly in yeast. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by

the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Fusion proteins provide an alternative for yeast expression systems, as well as in mammalian, baculovirus, and bacterial expression systems. Usually, a DNA sequence encoding the N-terminal portion of an endogenous yeast protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the yeast or human superoxide dismutase (SOD) gene, can be linked at the 5' terminus of a foreign gene and expressed in yeast. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. See *eg.* EP-A-0 196 056. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*eg.* ubiquitin-specific processing protease) to cleave the ubiquitin from the foreign protein. Through this method, therefore, native foreign protein can be isolated (*eg.* WO88/024066).

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provide for secretion in yeast of the foreign protein. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell.

DNA encoding suitable signal sequences can be derived from genes for secreted yeast proteins, such as the yeast invertase gene (EP-A-0 012 873; JPO. 62,096,086) and the A-factor gene (US patent 4,588,684). Alternatively, leaders of non-yeast origin, such as an interferon leader, exist that also provide for secretion in yeast (EP-A-0 060 057).

A preferred class of secretion leaders are those that employ a fragment of the yeast alpha-factor gene, which contains both a "pre" signal sequence, and a "pro" region. The types of alpha-factor fragments that can be employed include the full-length pre-pro alpha factor leader (about 83 amino acid residues) as well as truncated alpha-factor leaders (usually about 25 to about 50 amino acid residues) (US Patents 4,546,083 and 4,870,008; EP-A-0 324 274). Additional leaders employing an alpha-factor leader fragment that provides for secretion include hybrid alpha-factor leaders made with a presequence of a first yeast, but a pro-region from a second yeast alphafactor. (*eg.* see WO 89/02463.)

Usually, transcription termination sequences recognized by yeast are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator sequence and other yeast-recognized termination sequences, such as those coding for glycolytic enzymes.

Usually, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (*eg.* plasmids) capable of stable maintenance in a host, such as yeast or bacteria. The replicon may have two replication systems, thus allowing it to be maintained, for example, in yeast for expression and in a prokaryotic host for cloning and amplification. Examples of such yeast-bacteria shuttle vectors include YEp24 [Botstein *et al.* (1979) *Gene 8*:17-24], pCl/1 [Brake *et al.* (1984) *Proc. Natl. Acad. Sci USA 81*:4642-4646], and YRp17 [Stinchcomb *et al.* (1982) *J. Mol. Biol. 158*:157]. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably have at least about 10, and more preferably at least about 20. Enter a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host. See *eg.* Brake *et al.*, *supra*.

Alternatively, the expression constructs can be integrated into the yeast genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to a yeast chromosome that allows the vector to integrate, and preferably contain two homologous sequences flanking the expression construct. Integrations appear to result from recombinations between homologous DNA in the vector and the yeast chromosome [Orr-Weaver *et al.* (1983) *Methods in Enzymol. 101*:228-245]. An integrating vector may be directed to a specific locus in yeast by selecting the appropriate homologous sequence for inclusion in the vector. See Orr-Weaver *et al.*, *supra*. One or more expression construct may integrate, possibly affecting levels of recombinant protein produced [Rine *et al.* (1983) *Proc. Natl. Acad. Sci. USA 80*:6750]. The chromosomal sequences included in the vector can occur either as a single segment in the vector, which results in the integration of the entire vector, or two segments homologous to adjacent segments in the chromosome and flanking the expression construct in the vector, which can result in the stable integration of only the expression construct.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of yeast strains that have been transformed. Selectable markers may include biosynthetic genes that can be expressed in the yeast host, such as *ADE2*, *HIS4*, *LEU2*, *TRP1*, and *ALG7*, and the G418 resistance gene, which confer resistance in yeast cells to tunicamycin and G418, respectively. In addition, a suitable selectable marker may also provide yeast with the ability to grow in the presence of toxic compounds, such as metal. For example, the presence of *CUP1* allows yeast to grow in the presence of copper ions [Butt *et al.* (1987) *Microbiol, Rev. 51*:351].

Alternatively, some of the above described components can be put together into transformation vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeasts. For example, expression vectors have been developed for, *inter alia*, the following yeasts:Candida albicans [Kurtz, *et al.* (1986) *Mol. Cell. Biol. 6*:142], Candida maltosa [Kunze, *et al.* (1985) *J. Basic Microbiol. 25*:141]. Hansenula polymorpha [Gleeson, *et al.* (1986) *J. Gen. Microbiol. 132*:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet. 202*:302], Kluyveromyces fragilis [Das, *et al.* (1984) *J. Bacteriol. 158*:1165], Kluyveromyces lactis [De Louvencourt *et al.* (1983) *J. Bacteriol. 154*:737; Van den Berg *et al.* (1990) *Bio/Technology 8*:135], Pichia guillerimondii [Kunze *et al.* (1985) *J. Basic Microbiol. 25*:141], Pichia pastoris [Cregg, *et al.* (1985) *Mol. Cell. Biol. 5*:3376; US Patent Nos. 4,837,148 and 4,929,555], Saccharomyces cerevisiae [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA 75*:1929; Ito *et al.* (1983) *J. Bacteriol. 153*:163], Schizosaccharomyces pombe [Beach and Nurse (1981) *Nature 300*:706], and Yarrowia lipolytica [Davidow, *et al.* (1985) *Curr. Genet. 10*:380471 Gaillardin, *et al.* (1985) *Curr. Genet. 10*:49].

Methods of introducing exogenous DNA into yeast hosts are well-known in the art, and usually include either the transformation of spheroplasts or of intact yeast cells treated with alkali cations. Transformation procedures usually vary with the yeast species to be transformed. See *eg.* [Kurtz *et al.* (1986) *Mol. Cell. Biol. 6*:142; Kunze *et al.* (1985) *J. Basic Microbiol. 25*:141; Candida]; [Gleeson *et al.* (1986) *J. Gen. Microbiol. 132*:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet. 202*:302; Hansenula]; [Das *et al.* (1984) *J. Bacteriol. 158*:1165; De Louvencourt *et al.* (1983) *J. Bacteriol. 154*:1165; Van den Berg *et al.* (1990) *Bio/Technology 8*:135; Kluyveromyces]; [Cregg *et al.* (1985) *Mol. Cell. Biol. 5*:3376; Kunze *et al.* (1985) *J. Basic Microbiol. 25*:141; US Patent Nos. 4,837,148 and 4,929,555; Pichia]; [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA 75*;1929; Ito *et al.* (1983) *J. Bacteriol. 153*:163 Saccharomyces]; [Beach and Nurse (1981) *Nature 300*:706; Schizosaccharomyces]; [Davidow *et al.* (1985) *Curr. Genet. 10*:39; Gaillardin *et al.* (1985) *Curr. Genet. 10*:49; Yarrowia].

Antibodies

As used herein, the term "antibody" refers to a polypeptide or group of polypeptides composed of at least one antibody combining site. An "antibody combining site" is the three-dimensional binding space with an internal surface shape and charge distribution complementary to the features of an epitope of an antigen, which allows a binding of the antibody with the antigen. "Antibody" includes, for example, vertebrate antibodies, hybrid antibodies, chimeric antibodies, humanised antibodies, altered antibodies, univalent antibodies, Fab proteins, and single domain antibodies.

Antibodies against the proteins of the invention are useful for affinity chromatography, immunoassays, and distinguishing/identifying streptococcus proteins.

Antibodies to the proteins of the invention, both polyclonal and monoclonal, may be prepared by conventional methods. In general, the protein is first used to immunize a suitable animal, preferably a mouse, rat, rabbit or goat. Rabbits and goats are preferred for the preparation of polyclonal sera due to the volume of serum obtainable, and the availability of labeled anti-rabbit and anti-goat antibodies. Immunization is generally performed by mixing or emulsifying the protein in saline, preferably in an adjuvant such as Freund's complete adjuvant, and injecting the mixture or emulsion parenterally (generally subcutaneously or intramuscularly). A dose of 50-200 μg/injection is typically sufficient. Immunization is generally boosted 2-6 weeks later with one or more injections of the protein in saline, preferably using Freund's incomplete adjuvant. One may alternatively generate antibodies by in vitro immunization using methods known in the art, which for the purposes of this invention is considered equivalent to *in vivo* immunization. Polyclonal antisera is obtained by bleeding the immunized animal into a glass or plastic container, incubating the blood at 25°C for one hour, followed by incubating at 4°C for 2-18 hours. The serum is recovered by centrifugation (*eg.* 1,000*g* for 10 minutes). About 20-50 ml per bleed may be obtained from rabbits.

Monoclonal antibodies are prepared using the standard method of Kohler & Milstein [*Nature* (1975) 256:495-96], or a modification thereof. Typically, a mouse or rat is immunized as described above. However, rather than bleeding the animal to extract serum, the spleen (and optionally several large lymph nodes) is removed and dissociated into single cells. If desired, the

spleen cells may be screened (after removal of nonspecifically adherent cells) by applying a cell suspension to a plate or well coated with the protein antigen. B-cells expressing membrane-bound immunoglobulin specific for the antigen bind to the plate, and are not rinsed away with the rest of the suspension. Resulting B-cells, or all dissociated spleen cells, are then induced to fuse with myeloma cells to form hybridomas, and are cultured in a selective medium (*eg.* hypoxanthine, aminopterin, thymidine medium, "HAT"). The resulting hybridomas are plated by limiting dilution, and are assayed for production of antibodies which bind specifically to the immunizing antigen (and which do not bind to unrelated antigens). The selected MAb-secreting hybridomas are then cultured either *in vitro* (*eg.* in tissue culture bottles or hollow fiber reactors), or *in vivo* (as ascites in mice).

If desired, the antibodies (whether polyclonal or monoclonal) may be labeled using conventional techniques. Suitable labels include fluorophores, chromophores, radioactive atoms (particularly ^{32}P and ^{125}I), electron-dense reagents, enzymes, and ligands having specific binding partners. Enzymes are typically detected by their activity. For example, horseradish peroxidase is usually detected by its ability to convert 3,3',5,5'-tetramethylbenzidine (TMB) to a blue pigment, quantifiable with a spectrophotometer. "Specific binding partner" refers to a protein capable of binding a ligand molecule with high specificity, as for example in the case of an antigen and a monoclonal antibody specific therefor. Other specific binding partners include biotin and avidin or streptavidin, IgG and protein A, and the numerous receptor-ligand couples known in the art. It should be understood that the above description is not meant to categorize the various labels into distinct classes, as the same label may serve in several different modes. For example, ^{125}I may serve as a radioactive label or as an electron-dense reagent. HRP may serve as enzyme or as antigen for a MAb. Further, one may combine various labels for desired effect. For example, MAbs and avidin also require labels in the practice of this invention: thus, one might label a MAb with biotin, and detect its presence with avidin labeled with ^{125}I, or with an anti-biotin MAb labeled with HRP. Other permutations and possibilities will be readily apparent to those of ordinary skill in the art, and are considered as equivalents within the scope of the instant invention.

Pharmaceutical Compositions

Pharmaceutical compositions can comprise either polypeptides, antibodies, or nucleic acid of the invention. The pharmaceutical compositions will comprise a therapeutically effective amount of either polypeptides, antibodies, or polynucleotides of the claimed invention.

The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature. The precise effective amount for a subject will depend upon the subject's size and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount in advance. However, the effective amount for a given situation can be determined by routine experimentation and is within the judgement of the clinician.

For purposes of the present invention, an effective dose will be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the molecule of the invention in the individual to which it is administered.

A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Such carriers are well known to those of ordinary skill in the art.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles. Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier.

Delivery Methods

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (*eg.* see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Vaccines

Vaccines according to the invention may either be prophylactic (*ie.* to prevent infection) or therapeutic (*ie.* to treat disease after infection).

Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with "pharmaceutically acceptable carriers," which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, *etc.* pathogens.

Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to: (1) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59™ (WO90/14837; Chapter 10 in *Vaccine Design – the subunit and adjuvant approach* (1995) ed. Powell & Newman), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing MTP-PE) formulated into submicron particles using a microfluidizer, (b) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); (2) saponin adjuvants, such as QS21 or Stimulon™ (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes), which ISCOMS may be devoid of additional detergent *e.g.* WO00/07621; (3) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (4) cytokines, such as interleukins (*e.g.* IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12 (WO99/44636), *etc.*), interferons (*e.g.* gamma interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), *etc.*; (5) monophosphoryl lipid A (MPL) or 3-O-deacylated MPL (3dMPL) *e.g.* GB-2220221, EP-A-0689454; (6) combinations of 3dMPL with, for example, QS21 and/or oil-in-water emulsions *e.g.* EP-A-0835318, EP-A-0735898, EP-A-0761231; (7) oligonucleotides comprising CpG motifs [Krieg *Vaccine* 2000, 19, 618-622; Krieg *Curr opin Mol Ther* 2001 3:15-24; Roman *et al.*, *Nat. Med.*, 1997, 3, 849-854; Weiner *et al.*, *PNAS USA*, 1997, 94, 10833-10837; Davis *et al.*, *J. Immunol.*, 1998, 160, 870-876; Chu *et al.*, *J. Exp. Med.*, 1997, 186, 1623-1631; Lipford *et al.*, *Eur. J. Immunol.*, 1997, 27, 2340-2344; Moldoveanu *et al.*, *Vaccine*, 1988, 16, 1216-1224, Krieg *et al.*, *Nature*, 1995, 374, 546-549; Klinman *et al.*, *PNAS USA*, 1996, 93, 2879-2883; Ballas *et al.*, *J. Immunol.*, 1996, 157, 1840-1845; Cowdery *et al.*, *J. Immunol.*, 1996, 156, 4570-4575; Halpern *et al.*, *Cell. Immunol.*, 1996, 167, 72-78; Yamamoto *et al.*, *Jpn. J. Cancer Res.*, 1988, 79, 866-873; Stacey *et al.*, *J. Immunol.*, 1996, 157, 2116-2122; Messina *et al.*, *J. Immunol.*, 1991, 147, 1759-1764; Yi *et al.*, *J. Immunol.*, 1996, 157, 4918-4925; Yi *et al.*, *J. Immunol.*, 1996, 157, 5394-5402; Yi *et al.*, *J. Immunol.*, 1998, 160, 4755-4761; and Yi *et al.*, *J. Immunol.*, 1998, 160, 5898-5906; International patent applications WO96/02555, WO98/16247, WO98/18810, WO98/40100, WO98/55495, WO98/37919 and WO98/52581] *i.e.* containing at least one CG dinucleotide, with 5-methylcytosine optionally being used in place of cytosine; (8) a polyoxyethylene ether or a polyoxyethylene ester *e.g.* WO99/52549; (9) a polyoxyethylene sorbitan ester surfactant in combination with an octoxynol (*e.g.* WO01/21207) or a polyoxyethylene alkyl ether or ester surfactant in combination with at least one additional non-ionic surfactant such as an octoxynol (*e.g.* WO01/21152); (10) an immunostimulatory oligonucleotide (*e.g.* a CpG oligonucleotide) and a saponin *e.g.* WO00/62800; (11) an immunostimulant and a particle of metal salt *e.g.* WO00/23105; (12) a saponin and an oil-in-water emulsion *e.g.* WO99/11241; (13) a saponin (*e.g.* QS21) + 3dMPL + IL-12 (optionally + a sterol) *e.g.* WO98/57659; (14) aluminium salts, preferably hydroxide or phosphate, but any other suitable salt may also be used (*e.g.* hydroxyphosphate, oxyhydroxide, orthophosphate, sulphate *etc.* [*e.g.* see chapters 8 & 9 of Powell & Newman]). Mixtures of different aluminium

salts may also be used. The salt may take any suitable form (*e.g.* gel, crystalline, amorphous *etc.*); (15) other substances that act as immunostimulating agents to enhance the efficacy of the composition. Aluminium salts and/or MF59™ are preferred.

As mentioned above, muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-*sn*-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE), *etc.*

The immunogenic compositions (*eg.* the immunising antigen/immunogen/polypeptide/protein/ nucleic acid, pharmaceutically acceptable carrier, and adjuvant) typically will contain diluents, such as water, saline, glycerol, ethanol, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles.

Typically, the immunogenic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. The preparation also may be emulsified or encapsulated in liposomes for enhanced adjuvant effect, as discussed above under pharmaceutically acceptable carriers.

Immunogenic compositions used as vaccines comprise an immunologically effective amount of the antigenic or immunogenic polypeptides, as well as any other of the above-mentioned components, as needed. By "immunologically effective amount", it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention. This amount varies depending upon the health and physical condition of the individual to be treated, the taxonomic group of individual to be treated (*eg.* nonhuman primate, primate, *etc.*), the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials.

The immunogenic compositions are conventionally administered parenterally, *eg.* by injection, either subcutaneously, intramuscularly, or transdermally/transcutaneously (*eg.* WO98/20734). Additional formulations suitable for other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. The vaccine may be administered in conjunction with other immunoregulatory agents.

As an alternative to protein-based vaccines, DNA vaccination may be used [*eg.* Robinson & Torres (1997) *Seminars in Immunol* 9:271-283; Donnelly *et al.* (1997) *Annu Rev Immunol* 15:617-648; later herein].

Gene Delivery Vehicles

Gene therapy vehicles for delivery of constructs including a coding sequence of a therapeutic of the invention, to be delivered to the mammal for expression in the mammal, can be administered either locally or systemically. These constructs can utilize viral or non-viral vector approaches in *in vivo* or *ex vivo* modality. Expression of such coding sequence can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence in vivo can be either constitutive or regulated.

The invention includes gene delivery vehicles capable of expressing the contemplated nucleic acid sequences. The gene delivery vehicle is preferably a viral vector and, more preferably, a retroviral, adenoviral, adeno-associated viral (AAV), herpes viral, or alphavirus vector. The viral vector can also be an astrovirus, coronavirus, orthomyxovirus, papovavirus, paramyxovirus, parvovirus, picornavirus, poxvirus, or togavirus viral vector. See generally, Jolly (1994) *Cancer Gene Therapy* 1:51-64; Kimura (1994) *Human Gene Therapy* 5:845-852; Connelly (1995) *Human Gene Therapy* 6:185-193; and Kaplitt (1994) *Nature Genetics* 6:148-153.

Retroviral vectors are well known in the art and we contemplate that any retroviral gene therapy vector is employable in the invention, including B, C and D type retroviruses, xenotropic retroviruses (for example, NZB-X1, NZB-X2 and NZB9-1 (see O'Neill (1985) *J. Virol.* 53:160) polytropic retroviruses *eg.* MCF and MCF-MLV (see Kelly (1983) *J. Virol.* 45:291), spumaviruses and lentiviruses. See RNA Tumor Viruses, Second Edition, Cold Spring Harbor Laboratory, 1985.

Portions of the retroviral gene therapy vector may be derived from different retroviruses. For example, retrovector LTRs may be derived from a Murine Sarcoma Virus, a tRNA binding site from a Rous Sarcoma Virus, a packaging signal from a Murine Leukemia Virus, and an origin of second strand synthesis from an Avian Leukosis Virus.

These recombinant retroviral vectors may be used to generate transduction competent retroviral vector particles by introducing them into appropriate packaging cell lines (see US patent 5,591,624). Retrovirus vectors can be constructed for site-specific integration into host cell DNA by incorporation of a chimeric integrase enzyme into the retroviral particle (see WO96/37626). It is preferable that the recombinant viral vector is a replication defective recombinant virus.

Packaging cell lines suitable for use with the above-described retrovirus vectors are well known in the art, are readily prepared (see WO95/30763 and WO92/05266), and can be used to create producer cell lines (also termed vector cell lines or "VCLs") for the production of recombinant vector particles. Preferably, the packaging cell lines are made from human parent cells (*eg.* HT1080 cells) or mink parent cell lines, which eliminates inactivation in human serum.

Preferred retroviruses for the construction of retroviral gene therapy vectors include Avian Leukosis Virus, Bovine Leukemia, Virus, Murine Leukemia Virus, Mink-Cell Focus-Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis Virus and Rous Sarcoma Virus. Particularly preferred Murine Leukemia Viruses include 4070A and 1504A (Hartley and Rowe (1976) *J Virol* 19:19-25), Abelson (ATCC No. VR-999), Friend (ATCC No. VR-245), Graffi, Gross (ATCC Nol VR-590), Kirsten, Harvey Sarcoma Virus and Rauscher (ATCC No. VR-998) and Moloney Murine Leukemia Virus (ATCC No. VR-190). Such retroviruses may be obtained from depositories or collections such as the American Type Culture Collection ("ATCC") in Rockville, Maryland or isolated from known sources using commonly available techniques.

Exemplary known retroviral gene therapy vectors employable in this invention include those described in patent applications GB2200651, EP0415731, EP0345242, EP0334301, WO89/02468; WO89/05349, WO89/09271, WO90/02806, WO90/07936, WO94/03622, WO93/25698, WO93/25234, WO93/11230, WO93/10218, WO91/02805, WO91/02825, WO95/07994, US 5,219,740, US 4,405,712, US 4,861,719, US 4,980,289, US 4,777,127, US 5,591,624. See also Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res* 53:962-967; Ram (1993) *Cancer Res* 53 (1993) 83-88; Takamiya (1992) *J Neurosci Res* 33:493-503; Baba (1993) *J Neurosurg* 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *Proc Natl Acad Sci* 81:6349; and Miller (1990) *Human Gene Therapy* 1.

Human adenoviral gene therapy vectors are also known in the art and employable in this invention. See, for example, Berkner (1988) *Biotechniques* 6:616 and Rosenfeld (1991) *Science* 252:431, and WO93/07283, WO93/06223, and WO93/07282. Exemplary known adenoviral gene therapy vectors employable in this invention include those described in the above referenced documents and in WO94/12649, WO93/03769, WO93/19191, WO94/28938, WO95/11984, WO95/00655, WO95/27071, WO95/29993, WO95/34671, WO96/05320, WO94/08026, WO94/11506, WO93/06223, WO94/24299, WO95/14102, WO95/24297, WO95/02697, WO94/28152, WO94/24299, WO95/09241, WO95/25807, WO95/05835, WO94/18922 and WO95/09654. Alternatively, administration of DNA linked to killed adenovirus as described in Curiel (1992) *Hum. Gene Ther.* 3:147-154 may be employed. The gene delivery vehicles of the invention also include adenovirus associated virus (AAV) vectors. Leading and preferred examples of such vectors for use in this invention are the AAV-2 based vectors disclosed in Srivastava, WO93/09239. Most preferred AAV vectors comprise the two AAV inverted terminal repeats in which the native D-sequences are modified by substitution of nucleotides, such that at least 5 native nucleotides and up to 18 native nucleotides, preferably at least 10 native nucleotides up to 18 native nucleotides, most preferably 10 native nucleotides are retained and the remaining nucleotides of the D-sequence are deleted or replaced with non-native nucleotides. The native D-sequences of the AAV inverted terminal repeats are sequences of 20 consecutive nucleotides in each AAV inverted terminal repeat (*ie.* there is one sequence at each end) which are not involved in HP formation. The non-native replacement nucleotide may be any nucleotide other than the nucleotide found in the native D-sequence in the same position. Other employable exemplary AAV vectors are pWP-19, pWN-1, both of which are disclosed in Nahreini (1993) *Gene* 124:257-262. Another example of such an AAV vector is psub201 (see Samulski (1987) *J. Virol.* 61:3096). Another exemplary AAV vector is the Double-D ITR vector. Construction of the Double-D ITR vector is disclosed in US Patent 5,478,745. Still other vectors are those disclosed in Carter US Patent 4,797,368 and Muzyczka US Patent 5,139,941, Chartejee US Patent 5,474,935, and Kotin WO94/288157. Yet a further example of an AAV vector employable in this invention is SSV9AFABTKneo, which contains the AFP enhancer and albumin promoter and directs expression predominantly in the liver. Its structure and construction are disclosed in Su (1996) *Human Gene Therapy* 7:463-470. Additional AAV gene therapy vectors are described in US 5,354,678, US 5,173,414, US 5,139,941, and US 5,252,479.

The gene therapy vectors of the invention also include herpes vectors. Leading and preferred examples are herpes simplex virus vectors containing a sequence encoding a thymidine kinase polypeptide such as those disclosed in US 5,288,641 and EP0176170 (Roizman). Additional exemplary herpes simplex virus vectors include HFEM/ICP6-LacZ disclosed in WO95/04139 (Wistar Institute), pHSVlac described in Geller (1988) *Science* 241:1667-1669 and in WO90/09441 and WO92/07945, HSV Us3::pgC-lacZ described in Fink (1992) *Human Gene Therapy* 3:11-19 and HSV 7134, 2 RH 105 and GAL4 described in EP 0453242 (Breakefield), and those deposited with the ATCC with accession numbers VR-977 and VR-260.

Also contemplated are alpha virus gene therapy vectors that can be employed in this invention. Preferred alpha virus vectors are Sindbis viruses vectors. Togaviruses, Semliki Forest virus (ATCC VR-67; ATCC VR-1247), Middleberg virus (ATCC

VR-370), Ross River virus (ATCC VR-373; ATCC VR-1246), Venezuelan equine encephalitis virus (ATCC VR923; ATCC VR-1250; ATCC VR-1249; ATCC VR-532), and those described in US patents 5,091,309, 5,217,879, and WO92/10578. More particularly, those alpha virus vectors described in US Serial No. 08/405,627, filed March 15, 1995,WO94/21792, WO92/10578, WO95/07994, US 5,091,309 and US 5,217,879 are employable. Such alpha viruses may be obtained from depositories or collections such as the ATCC in Rockville, Maryland or isolated from known sources using commonly available techniques. Preferably, alphavirus vectors with reduced cytotoxicity are used (see USSN 08/679640).

DNA vector systems such as eukaryotic layered expression systems are also useful for expressing the nucleic acids of the invention. See WO95/07994 for a detailed description of eukaryotic layered expression systems. Preferably, the eukaryotic layered expression systems of the invention are derived from alphavirus vectors and most preferably from Sindbis viral vectors.

Other viral vectors suitable for use in the present invention include those derived from poliovirus, for example ATCC VR-58 and those described in Evans, Nature 339 (1989) 385 and Sabin (1973) *J. Biol. Standardization* 1:115; rhinovirus, for example ATCC VR-1110 and those described in Arnold (1990) *J Cell Biochem* L401; pox viruses such as canary pox virus or vaccinia virus, for example ATCC VR-111 and ATCC VR-2010 and those described in Fisher-Hoch (1989) *Proc Natl Acad Sci* 86:317; Flexner (1989) *Ann NY Acad Sci* 569:86, Flexner (1990) *Vaccine* 8:17; in US 4,603,112 and US 4,769,330 and WO89/01973; SV40 virus, for example ATCC VR-305 and those described in Mulligan (1979) *Nature* 277:108 and Madzak (1992) *J Gen Virol* 73:1533; influenza virus, for example ATCC VR-797 and recombinant influenza viruses made employing reverse genetics techniques as described in US 5,166,057 and in Enami (1990) *Proc Natl Acad Sci* 87:3802-3805; Enami & Palese (1991) *J Virol* 65:2711-2713 and Luytjes (1989) *Cell* 59:110, (see also McMichael (1983) *NEJ Med* 309:13, and Yap (1978) *Nature* 273:238 and *Nature* (1979) 277:108); human immunodeficiency virus as described in EP-0386882 and in Buchschacher (1992) *J. Virol.* 66:2731; measles virus, for example ATCC VR-67 and VR-1247 and those described in EP-0440219; Aura virus, for example ATCC VR-368; Bebaru virus, for example ATCC VR-600 and ATCC VR-1240; Cabassou virus, for example ATCC VR-922; Chikungunya virus, for example ATCC VR-64 and ATCC VR-1241; Fort Morgan Virus, for example ATCC VR-924; Getah virus, for example ATCC VR-369 and ATCC VR-1243; Kyzylagach virus, for example ATCC VR-927; Mayaro virus, for example ATCC VR-66; Mucambo virus, for example ATCC VR-580 and ATCC VR-1244; Ndumu virus, for example ATCC VR-371; Pixuna virus, for example ATCC VR-372 and ATCC VR-1245; Tonate virus, for example ATCC VR-925; Triniti virus, for example ATCC VR-469; Una virus, for example ATCC VR-374; Whataroa virus, for example ATCC VR-926; Y-62-33 virus, for example ATCC VR-375; O'Nyong virus, Eastern encephalitis virus, for example ATCC VR-65 and ATCC VR-1242; Western encephalitis virus, for example ATCC VR-70, ATCC VR-1251, ATCC VR-622 and ATCC VR-1252; and coronavirus, for example ATCC VR-740 and those described in Hamre (1966) *Proc Soc Exp Biol Med* 121:190.

Delivery of the compositions of this invention into cells is not limited to the above mentioned viral vectors. Other delivery methods and media may be employed such as, for example, nucleic acid expression vectors, polycationic condensed DNA linked or unlinked to killed adenovirus alone, for example see US Serial No. 08/366,787, filed December 30, 1994 and Curiel (1992) *Hum Gene Ther* 3:147-154 ligand linked DNA, for example see Wu (1989) *J Biol Chem* 264:16985-16987, eucaryotic cell delivery vehicles cells, for example see US Serial No.08/240,030, filed May 9, 1994, and US Serial No. 08/404,796, deposition of photopolymerized hydrogel materials, hand-held gene transfer particle gun, as described in US Patent 5,149,655, ionizing radiation as described in US5,206,152 and in WO92/11033, nucleic charge neutralization or fusion with cell membranes. Additional approaches are described in Philip (1994) *Mol Cell Biol* 14:2411-2418 and in Woffendin (1994) *Proc Natl Acad Sci* 91:1581-1585.

Particle mediated gene transfer may be employed, for example see US Serial No. 60/023,867. Briefly, the sequence can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, as described in Wu & Wu (1987) *J. Biol. Chem.* 262:4429-4432, insulin as described in Hucked (1990) *Biochem Pharmacol* 40:253-263, galactose as described in Plank (1992) *Bioconjugate Chem* 3:533-539, lactose or transferrin.

Naked DNA may also be employed. Exemplary naked DNA introduction methods are described in WO 90/11092 and US 5,580,859. Uptake efficiency may be improved using biodegradable latex beads. DNA coated latex beads are efficiently transported into cells after endocytosis initiation by the beads. The method may be improved further by treatment of the beads to increase hydrophobicity and thereby facilitate disruption of the endosome and release of the DNA into the cytoplasm.

Liposomes that can act as gene delivery vehicles are described in US 5,422,120, WO95/13796, WO94/23697, WO91/14445 and EP-524,968. As described in USSN. 60/023,867, on non-viral delivery, the nucleic acid sequences encoding a polypeptide

can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then be incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, insulin, galactose, lactose, or transferrin. Other delivery systems include the use of liposomes to encapsulate DNA comprising the gene under the control of a variety of tissue-specific or ubiquitously-active promoters. Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin *et al* (1994) *Proc. Natl. Acad. Sci. USA* 91(24):11581-11585. Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials. Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun, as described in US 5,149,655; use of ionizing radiation for activating transferred gene, as described in US 5,206,152 and WO92/11033

Exemplary liposome and polycationic gene delivery vehicles are those described in US 5,422,120 and 4,762,915; in WO 95/13796; WO94/23697; and WO91/14445; in EP-0524968; and in Stryer, Biochemistry, pages 236-240 (1975) W.H. Freeman, San Francisco; Szoka (1980) *Biochem Biophys Acta* 600:1; Bayer (1979) *Biochem Biophys Acta* 550:464; Rivnay (1987) *Meth Enzymol* 149:119; Wang (1987) *Proc Natl Acad Sci* 84:7851; Plant (1989) *Anal Biochem* 176:420.

A polynucleotide composition can comprises therapeutically effective amount of a gene therapy vehicle, as the term is defined above. For purposes of the present invention, an effective dose will be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

Delivery Methods

Once formulated, the polynucleotide compositions of the invention can be administered (1) directly to the subject; (2) delivered *ex vivo*, to cells derived from the subject; or (3) *in vitro* for expression of recombinant proteins. The subjects to be treated can be mammals or birds. Also, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (*eg.* see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art and described in *eg.* WO93/14778. Examples of cells useful in ex vivo applications include, for example, stem cells, particularly hematopoetic, lymph cells, macrophages, dendritic cells, or tumor cells.

Generally, delivery of nucleic acids for both *ex vivo* and *in vitro* applications can be accomplished by the following procedures, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.

Polynucleotide and polypeptide pharmaceutical compositions

In addition to the pharmaceutically acceptable carriers and salts described above, the following additional agents can be used with polynucleotide and/or polypeptide compositions.

A.Polypeptides

One example are polypeptides which include, without limitation: asioloorosomucoid (ASOR); transferrin; asialoglycoproteins; antibodies; antibody fragments; ferritin; interleukins; interferons, granulocyte, macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), macrophage colony stimulating factor (M-CSF), stem cell factor and erythropoietin. Viral antigens, such as envelope proteins, can also be used. Also, proteins from other invasive organisms, such as the 17 amino acid peptide from the circumsporozoite protein of plasmodium falciparum known as RII.

B.Hormones, Vitamins, *etc.*

Other groups that can be included are, for example: hormones, steroids, androgens, estrogens, thyroid hormone, or vitamins, folic acid.

C.Polyalkylenes, Polysaccharides, *etc.*

Also, polyalkylene glycol can be included with the desired polynucleotides/polypeptides. In a preferred embodiment, the polyalkylene glycol is polyethlylene glycol. In addition, mono-, di-, or polysaccharides can be included. In a preferred embodiment of this aspect, the polysaccharide is dextran or DEAE-dextran. Also, chitosan and poly(lactide-co-glycolide)

D.Lipids, and Liposomes

The desired polynucleotide/polypeptide can also be encapsulated in lipids or packaged in liposomes prior to delivery to the subject or to cells derived therefrom.

Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed polynucleotide to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight (1991) *Biochim. Biophys. Acta.* 1097:1-17; Straubinger (1983) *Meth. Enzymol.* 101:512-527.

Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations. Cationic liposomes have been shown to mediate intracellular delivery of plasmid DNA (Felgner (1987) *Proc. Natl. Acad. Sci. USA* 84:7413-7416); mRNA (Malone (1989) *Proc. Natl. Acad. Sci. USA* 86:6077-6081); and purified transcription factors (Debs (1990) *J. Biol. Chem.* 265:10189-10192), in functional form.

Cationic liposomes are readily available. For example, N[1-2,3-dioleyloxy)propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. (See, also, Felgner *supra*). Other commercially available liposomes include transfectace (DDAB/DOPE) and DOTAP/DOPE (Boerhinger). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, *eg.* Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; WO90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

Similarly, anionic and neutral liposomes are readily available, such as from Avanti Polar Lipids (Birmingham, AL), or can be easily prepared using readily available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphoshatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the DOTMA and DOTAP starting materials in appropriate ratios. Methods for making liposomes using these materials are well known in the art.

The liposomes can comprise multilammelar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See *eg.* Straubinger (1983) *Meth. Immunol.* 101:512-527; Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; Papahadjopoulos (1975) *Biochim. Biophys. Acta* 394:483; Wilson (1979) *Cell* 17:77); Deamer & Bangham (1976) *Biochim. Biophys. Acta* 443:629; Ostro (1977) *Biochem. Biophys. Res. Commun.* 76:836; Fraley (1979) *Proc. Natl. Acad. Sci. USA* 76:3348); Enoch & Strittmatter (1979) *Proc. Natl. Acad. Sci. USA* 76:145; Fraley (1980) *J. Biol. Chem.* (1980) 255:10431; Szoka & Papahadjopoulos (1978) *Proc. Natl. Acad. Sci. USA* 75:145; and Schaefer-Ridder (1982) *Science* 215:166.

E.Lipoproteins

In addition, lipoproteins can be included with the polynucleotide/polypeptide to be delivered. Examples of lipoproteins to be utilized include: chylomicrons, HDL, IDL, LDL, and VLDL. Mutants, fragments, or fusions of these proteins can also be used. Also, modifications of naturally occurring lipoproteins can be used, such as acetylated LDL. These lipoproteins can target the delivery of polynucleotides to cells expressing lipoprotein receptors. Preferably, if lipoproteins are including with the polynucleotide to be delivered, no other targeting ligand is included in the composition.

Naturally occurring lipoproteins comprise a lipid and a protein portion. The protein portion are known as apoproteins. At the present, apoproteins A, B, C, D, and E have been isolated and identified. At least two of these contain several proteins, designated by Roman numerals, AI, AII, AIV; CI, CII, CIII.

A lipoprotein can comprise more than one apoprotein. For example, naturally occurring chylomicrons comprises of A, B, C & E, over time these lipoproteins lose A and acquire C & E. VLDL comprises A, B, C & E apoproteins, LDL comprises apoprotein B; and HDL comprises apoproteins A, C, & E.

The amino acid of these apoproteins are known and are described in, for example, Breslow (1985) Annu Rev. Biochem 54:699; Law (1986) Adv. Exp Med. Biol. 151:162; Chen (1986) J Biol Chem 261:12918; Kane (1980) Proc Natl Acad Sci USA 77:2465; and Utermann (1984) Hum Genet 65:232.

Lipoproteins contain a variety of lipids including, triglycerides, cholesterol (free and esters), and phospholipids. The composition of the lipids varies in naturally occurring lipoproteins. For example, chylomicrons comprise mainly triglycerides. A more detailed description of the lipid content of naturally occurring lipoproteins can be found, for example, in *Meth. Enzymol.* 128 (1986). The composition of the lipids are chosen to aid in conformation of the apoprotein for receptor binding activity. The composition of lipids can also be chosen to facilitate hydrophobic interaction and association with the polynucleotide binding molecule.

Naturally occurring lipoproteins can be isolated from serum by ultracentrifugation, for instance. Such methods are described in *Meth. Enzymol.* (*supra)*; Pitas (1980) *J. Biochem.* 255:5454-5460 and Mahey (1979) *J Clin. Invest* 64:743-750. Lipoproteins can also be produced by *in vitro* or recombinant methods by expression of the apoprotein genes in a desired host cell. See, for example, Atkinson (1986) *Annu Rev Biophys Chem* 15:403 and Radding (1958) *Biochim Biophys Acta* 30: 443. Lipoproteins can also be purchased from commercial suppliers, such as Biomedical Techniologies, Inc., Stoughton, MA, USA. Further description of lipoproteins can be found in WO98/06437..

F.Polycationic Agents

Polycationic agents can be included, with or without lipoprotein, in a composition with the desired polynucleotide/polypeptide to be delivered.

Polycationic agents, typically, exhibit a net positive charge at physiological relevant pH and are capable of neutralizing the electrical charge of nucleic acids to facilitate delivery to a desired location. These agents have both in vitro, ex vivo, and in vivo applications. Polycationic agents can be used to deliver nucleic acids to a living subject either intramuscularly, subcutaneously, *etc.*

The following are examples of useful polypeptides as polycationic agents: polylysine, polyarginine, polyornithine, and protamine. Other examples include histones, protamines, human serum albumin, DNA binding proteins, non-histone chromosomal proteins, coat proteins from DNA viruses, such as (X174, transcriptional factors also contain domains that bind DNA and therefore may be useful as nucleic aid condensing agents. Briefly, transcriptional factors such as C/CEBP, c-jun, c-fos, AP-1, AP-2, AP-3, CPF, Prot-1, Sp-1, Oct-1, Oct-2, CREP, and TFIID contain basic domains that bind DNA sequences.

Organic polycationic agents include: spermine, spermidine, and purtrescine.

The dimensions and of the physical properties of a polycationic agent can be extrapolated from the list above, to construct other polypeptide polycationic agents or to produce synthetic polycationic agents.

Synthetic polycationic agents which are useful include, for example, DEAE-dextran, polybrene. Lipofectin™, and lipofectAMINE™ are monomers that form polycationic complexes when combined with polynucleotides/polypeptides.

Immunodiagnostic Assays

Streptococcus antigens of the invention can be used in immunoassays to detect antibody levels (or, conversely, anti-streptococcus antibodies can be used to detect antigen levels). Immunoassays based on well defined, recombinant antigens can be developed to replace invasive diagnostics methods. Antibodies to streptococcus proteins within biological samples, including for example, blood or serum samples, can be detected. Design of the immunoassays is subject to a great deal of variation, and a variety of these are known in the art. Protocols for the immunoassay may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also known; examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

Kits suitable for immunodiagnosis and containing the appropriate labeled reagents are constructed by packaging the appropriate materials, including the compositions of the invention, in suitable containers, along with the remaining reagents and materials (for example, suitable buffers, salt solutions, *etc.*) required for the conduct of the assay, as well as suitable set of assay instructions.

Nucleic Acid Hybridisation

"Hybridization" refers to the association of two nucleic acid sequences to one another by hydrogen bonding. Typically, one sequence will be fixed to a solid support and the other will be free in solution. Then, the two sequences will be placed in contact with one another under conditions that favor hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase sequence to the solid support (Denhardt's reagent or BLOTTO); concentration of the sequences; use of compounds to increase the rate of association of sequences (dextran sulfate or polyethylene glycol); and the stringency of the washing conditions following hybridization. See Sambrook *et al.* [*supra*] Volume 2, chapter 9, pages 9.47 to 9.57.

"Stringency" refers to conditions in a hybridization reaction that favor association of very similar sequences over sequences that differ. For example, the combination of temperature and salt concentration should be chosen that is approximately 120 to 200°C below the calculated Tm of the hybrid under study. The temperature and salt conditions can often be determined empirically in preliminary experiments in which samples of genomic DNA immobilized on filters are hybridized to the sequence of interest and then washed under conditions of different stringencies. See Sambrook *et al.* at page 9.50.

Variables to consider when performing, for example, a Southern blot are (1) the complexity of the DNA being blotted and (2) the homology between the probe and the sequences being detected. The total amount of the fragment(s) to be studied can vary a magnitude of 10, from 0.1 to 1µg for a plasmid or phage digest to 10^{-9} to 10^{-8} g for a single copy gene in a highly complex eukaryotic genome. For lower complexity polynucleotides, substantially shorter blotting, hybridization, and exposure times, a smaller amount of starting polynucleotides, and lower specific activity of probes can be used. For example, a single-copy yeast gene can be detected with an exposure time of only 1 hour starting with 1 µg of yeast DNA, blotting for two hours, and hybridizing for 4-8 hours with a probe of 10^8 cpm/µg. For a single-copy mammalian gene a conservative approach would start with 10 µg of DNA, blot overnight, and hybridize overnight in the presence of 10% dextran sulfate using a probe of greater than 10^8 cpm/µg, resulting in an exposure time of ~24 hours.

Several factors can affect the melting temperature (Tm) of a DNA-DNA hybrid between the probe and the fragment of interest, and consequently, the appropriate conditions for hybridization and washing. In many cases the probe is not 100% homologous to the fragment. Other commonly encountered variables include the length and total G+C content of the hybridizing sequences and the ionic strength and formamide content of the hybridization buffer. The effects of all of these factors can be approximated by a single equation:

$$Tm = 81 + 16.6(\log_{10}Ci) + 0.4[\%(G + C)] - 0.6(\%formamide) - 600/n - 1.5(\%mismatch).$$

where Ci is the salt concentration (monovalent ions) and n is the length of the hybrid in base pairs (slightly modified from Meinkoth & Wahl (1984) *Anal. Biochem.* 138: 267-284).

In designing a hybridization experiment, some factors affecting nucleic acid hybridization can be conveniently altered. The temperature of the hybridization and washes and the salt concentration during the washes are the simplest to adjust. As the temperature of the hybridization increases (*ie.* stringency), it becomes less likely for hybridization to occur between strands that are nonhomologous, and as a result, background decreases. If the radiolabeled probe is not completely homologous with the immobilized fragment (as is frequently the case in gene family and interspecies hybridization experiments), the hybridization temperature must be reduced, and background will increase. The temperature of the washes affects the intensity of the hybridizing band and the degree of background in a similar manner. The stringency of the washes is also increased with decreasing salt concentrations.

In general, convenient hybridization temperatures in the presence of 50% formamide are 42°C for a probe with is 95% to 100% homologous to the target fragment, 37°C for 90% to 95% homology, and 32°C for 85% to 90% homology. For lower homologies, formamide content should be lowered and temperature adjusted accordingly, using the equation above. If the homology between the probe and the target fragment are not known, the simplest approach is to start with both hybridization and wash conditions which are nonstringent. If non-specific bands or high background are observed after autoradiography, the filter can be washed at high stringency and reexposed. If the time required for exposure makes this approach impractical, several hybridization and/or washing stringencies should be tested in parallel.

Nucleic Acid Probe Assays

Methods such as PCR, branched DNA probe assays, or blotting techniques utilizing nucleic acid probes according to the invention can determine the presence of cDNA or mRNA. A probe is said to "hybridize" with a sequence of the invention if it can form a duplex or double stranded complex, which is stable enough to be detected.

The nucleic acid probes will hybridize to the streptococcus nucleotide sequences of the invention (including both sense and antisense strands). Though many different nucleotide sequences will encode the amino acid sequence, the native streptococcus sequence is preferred because it is the actual sequence present in cells. mRNA represents a coding sequence and so a probe should be complementary to the coding sequence; single-stranded cDNA is complementary to mRNA, and so a cDNA probe should be complementary to the non-coding sequence.

The probe sequence need not be identical to the streptococcus sequence (or its complement) — some variation in the sequence and length can lead to increased assay sensitivity if the nucleic acid probe can form a duplex with target nucleotides, which can be detected. Also, the nucleic acid probe can include additional nucleotides to stabilize the formed duplex. Additional streptococcus sequence may also be helpful as a label to detect the formed duplex. For example, a non-complementary nucleotide sequence

may be attached to the 5' end of the probe, with the remainder of the probe sequence being complementary to a streptococcus sequence. Alternatively, non-complementary bases or longer sequences can be interspersed into the probe, provided that the probe sequence has sufficient complementarity with the a streptococcus sequence in order to hybridize therewith and thereby form a duplex which can be detected.

The exact length and sequence of the probe will depend on the hybridization conditions (*e.g.* temperature, salt condition *etc.*). For example, for diagnostic applications, depending on the complexity of the analyte sequence, the nucleic acid probe typically contains at least 10-20 nucleotides, preferably 15-25, and more preferably at least 30 nucleotides, although it may be shorter than this. Short primers generally require cooler temperatures to form sufficiently stable hybrid complexes with the template.

Probes may be produced by synthetic procedures, such as the triester method of Matteucci *et al.* [*J. Am. Chem. Soc.* (1981) 103:3185], or according to Urdea *et al.* [*Proc. Natl. Acad. Sci. USA* (1983) 80: 7461], or using commercially available automated oligonucleotide synthesizers.

The chemical nature of the probe can be selected according to preference. For certain applications, DNA or RNA are appropriate. For other applications, modifications may be incorporated *eg.* backbone modifications, such as phosphorothioates or methylphosphonates, can be used to increase *in vivo* half-life, alter RNA affinity, increase nuclease resistance *etc.* [*eg.* see Agrawal & Iyer (1995) *Curr Opin Biotechnol* 6:12-19; Agrawal (1996) *TIBTECH* 14:376-387]; analogues such as peptide nucleic acids may also be used [*eg.* see Corey (1997) *TIBTECH* 15:224-229; Buchardt *et al.* (1993) *TIBTECH* 11:384-386].

Alternatively, the polymerase chain reaction (PCR) is another well-known means for detecting small amounts of target nucleic acid. The assay is described in Mullis *et al.* [*Meth. Enzymol.* (1987) 155:335-350] & US patents 4,683,195 & 4,683,202. Two "primer" nucleotides hybridize with the target nucleic acids and are used to prime the reaction. The primers can comprise sequence that does not hybridize to the sequence of the amplification target (or its complement) to aid with duplex stability or, for example, to incorporate a convenient restriction site. Typically, such sequence will flank the desired streptococcus sequence.

A thermostable polymerase creates copies of target nucleic acids from the primers using the original target nucleic acids as a template. After a threshold amount of target nucleic acids are generated by the polymerase, they can be detected by more traditional methods, such as Southern blots. When using the Southern blot method, the labelled probe will hybridize to the streptococcus sequence (or its complement).

Also, mRNA or cDNA can be detected by traditional blotting techniques described in Sambrook *et al* [*supra*]. mRNA, or cDNA generated from mRNA using a polymerase enzyme, can be purified and separated using gel electrophoresis. The nucleic acids on the gel are then blotted onto a solid support, such as nitrocellulose. The solid support is exposed to a labelled probe and then washed to remove any unhybridized probe. Next, the duplexes containing the labeled probe are detected. Typically, the probe is labelled with a radioactive moiety.

BRIEF DESCRIPTION OF DRAWINGS

Figures 1 to **85**, **119** to **188**, **238** and **239** show SDS-PAGE analysis of total cell extracts from cultures of recombinant *E.coli* expressing GBS proteins of the invention. Lane 1 in each gel (except for Figure 185) contains molecular weight markers. These are 94, 67, 43, 30, 20.1 & 14.4 kDa (except for Figures 7, 8, 10, 11, 13, 14, 15 and 119-170, which use 250, 150, 100, 75, 50, 37, 25, 15 & 10 kDa).

Figure 86A shows the pDEST15 vector and **Figure 86B** shows the pDEST17-1 vector.

Figures 88 to **118** and **247** to **319** show protein characterisation data for various proteins of the invention.

Figures 189 to **237** and **240** to **246** show SDS-PAGE analysis of purified GBS proteins of the invention. The left-hand lane contains molecular weight markers. These are 94, 67, 43, 30, 20.1 & 14.4 kDa.

MODES FOR CARRYING OUT THE INVENTION

The following examples describe nucleic acid sequences which have been identified in *Streptococcus*, along with their inferred translation products. The examples are generally in the following format:

- a nucleotide sequence which has been identified in *Streptococcus*
- the inferred translation product of this sequence
- a computer analysis (*e.g.* PSORT output) of the translation product, indicating antigenicity

Most examples describe nucleotide sequences from *S.agalactiae.* The specific strain which was sequenced was from serotype V, and is a clinical strain isolated in Italy which expresses the R antigen (ISS/Rome/Italy collection, strain.2603 V/R). For several of these examples, the corresponding sequences from *S.pyogenes* are also given. Where GBS and GAS show homology in this way, there is conservation between species which suggests an essential function and also gives good cross-species reactivity.

In contrast, several examples describe nucleotide sequences from GAS for which no homolog in GBS has been identified. This lack of homology gives molecules which are useful for distinguishing GAS from GBS and for making GAS-specific products. The same is true for GBS sequences which lack GAS homologs *e.g.* these are useful for making GBS-specific products.

The examples typically include details of homology to sequences in the public databases. Proteins that are similar in sequence are generally similar in both structure and function, and the homology often indicates a common evolutionary origin. Comparison with sequences of proteins of known function is widely used as a guide for the assignment of putative protein function to a new sequence and has proved particularly useful in whole-genome analyses.

Various tests can be used to assess the *in vivo* immunogenicity of the proteins identified in the examples. For example, the proteins can be expressed recombinantly and used to screen patient sera by immunoblot. A positive reaction between the protein and patient serum indicates that the patient has previously mounted an immune response to the protein in question *i.e.* the protein is an immunogen. This method can also be used to identify immunodominant proteins. The mouse model used in the examples can also be used.

The recombinant protein can also be conveniently used to prepare antibodies *e.g.* in a mouse. These can be used for direct confirmation that a protein is located on the cell-surface. Labelled antibody (*e.g.* fluorescent labelling for FACS) can be incubated with intact bacteria and the presence of label on the bacterial surface confirms the location of the protein.

For many GBS proteins, the following data are given:

- SDS-PAGE analysis of total recombinant *E.coli* cell extracts for GBS protein expression
- SDS-PAGE analysis after the protein purification

- Western-blot analysis of GBS total cell extract using antisera raised against recombinant proteins
- FACS and ELISA analysis against GBS using antisera raise against recombinant proteins
- Results of the *in vivo* passive protection assay

Details of experimental techniques used are presented below:

Sequence analysis

Open reading frames (ORFs) within nucleotide sequences were predicted using the GLIMMER program [Salzberg *et al.* (1998) *Nucleic Acids Res* 26:544-8]. Where necessary, start codons were modified and corrected manually on the basis of the presence of ribosome-binding sites and promoter regions on the upstream DNA sequence.

ORFs were then screened against the non-redundant protein databases using the programs BLASTp [Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-410] and PRAZE, a modification of the Smith-Waterman algorithm [Smith & Waterman (1981) *J Mol Biol* 147:195-7; see Fleischmann *et al* (1995) *Science* 269:496-512].

Leader peptides within the ORFs were located using three different approaches: (i) PSORT [Nakai (1991) *Bull. Inst. Chem. Res., Kyoto Univ.* 69:269-291; Horton & Nakai (1996) *Intellig. Syst. Mol. Biol.* 4:109-115; Horton & Nakai (1997) *Intellig. Syst. Mol. Biol.* 5:147-152]; (ii) SignalP [Nielsen & Krogh (1998) in *Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology* (*ISMB* 6), AAAI Press, Menlo Park, California, pp. 122-130; Nielsen *et al.* (1999) *Protein Engineering* 12:3-9; Nielsen *et al.* (1997). *Int. J. Neural Sys.* 8:581-599]; and (iii) visual inspection of the ORF sequences. Where a signal sequences is given a "possible site" value, the value represents the C-terminus residue of the signal peptide *e.g.* a "possible site" of 26 means that the signal sequence consists of amino acids 1-26.

Lipoprotein-specific signal peptides were located using three different approaches: (i) PSORT [see above]; (ii) the "prokaryotic membrane lipoprotein lipid attachment site" PROSITE motif [Hofmann *et al.* (1999) *Nucleic Acids Res.* 27:215-219; Bucher & Bairoch (1994) in *Proceedings 2nd International Conference on Intelligent Systems for Molecular Biology* (ISMB-94), AAAI Press, pages 53-61]; and (iii) the FINDPATTERNS program available in the GCG Wisconsin Package, using the pattern `(M,L,V)x{9,35}LxxCx`.

Transmembrane domains were located using two approaches: (i) PSORT [see above]; (ii) TopPred [von Heijne (1992) *J. Mol. Biol.* 225:487-494].

LPXTG motifs, characteristic of cell-wall attached proteins in Gram-positive bacteria [Fischetti *et al.* (1990) *Mol Microbiol* 4:1603-5] were located with FINDPATTERNS using the pattern `(L,I,V,M,Y,F)Px(T,A,S,G)(G,N,S,T,A,L)`.

RGD motifs, characteristic of cell-adhesion molecules [D'Souza *et al.* (1991) *Trends Biochem Sci* 16:246-50] were located using FINDPATTERNS.

Enzymes belonging to the glycolytic pathway were also selected as antigens, because these have been found experimentally expressed on the surface of *Streptococci* [*e.g.* Pancholi & Fischetti (1992) *J Exp Med* 176:415-26; Pancholi & Fischetti (1998) *J Biol Chem* 273:14503-15].

Cloning, expression and purification of proteins

GBS genes were cloned to facilitate expression in *E.coli* as two different types of fusion proteins:

a) proteins having a hexa-histidine tag at the amino-terminus (His-gbs)

b) proteins having a GST fusion partner at the amino-terminus (Gst-gbs)

Cloning was performed using the Gateway™ technology (Life Technologies), which is based on the site-specific recombination reactions that mediate integration and excision of phage lambda into and from the *E.coli* genome. A single cloning experiment included the following steps:

1- Amplification of GBS chromosomal DNA to obtain a PCR product coding for a single ORF flanked by *att*B recombination sites.

2- Insertion of the PCR product into a pDONR vector (containing *att*P sites) through a BP reaction (*att*B x *att*P sites). This reaction gives a so called 'pEntry' vector, which now contains *att*L sites flanking the insert.

3- Insertion of the GBS gene into *E.coli* expression vectors (pDestination vectors, containing *att*R sites) through a LR reaction between pEntry and pDestination plasmids (*att*L x *att*R sites).

A) Chromosomal DNA preparation

For chromosomal DNA preparation, GBS strain 2603 V/R (Istituto Superiore Sanità, Rome) was grown to exponential phase in 2 litres TH Broth (Difco) at 37°C, harvested by centrifugation, and dissolved in 40 ml TES (50 mM Tris pH 8, 5 mM EDTA pH 8, 20% sucrose). After addition of 2.5 ml lysozyme solution (25 mg/ml in TES) and 0.5 ml mutanolysin (Sigma M-9901, 25000U/ml in H_2O), the suspension was incubated at 37°C for 1 hour. 1 ml RNase (20 mg/ml) and 0.1 ml proteinase K (20 mg/ml) were added and incubation was continued for 30 min. at 37°C.

Cell lysis was obtained by adding 5 ml sarkosyl solution (10% N-laurylsarcosine in 250 mM EDTA pH 8.0), and incubating 1 hour at 37°C with frequent inversion. After sequential extraction with phenol, phenol-chloroform and chloroform, DNA was precipitated with 0.3M sodium acetate pH 5.2 and 2 volumes of absolute ethanol. The DNA pellet was rinsed with 70% ethanol and dissolved in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8). DNA concentration was evaluated by OD_{260}.

B) Oligonucleotide design

Synthetic oligonucleotide primers were designed on the basis of the coding sequence of each ORF. The aim was to express the protein's extracellular region. Accordingly, predicted signal peptides were omitted (by deducing the 5' end amplification primer sequence immediately downstream from the predicted leader sequence) and C-terminal cell-wall ancoring regions were removed (*e.g.* LPXTG motifs and downstream amino acids). Where additional nucleotides have been deleted, this is indicated by the suffix 'd' (*e.g.* 'GBS352d' – see Table V). Conversely, a suffix 'L' refers to expression without these deletions. Deletions of C- or N-terminal residues were also sometimes made, as indicated by a 'C' or 'N' suffix.

The amino acid sequences of the expressed GBS proteins (including 'd' and 'L' forms *etc.*) are definitively defined by the sequences of the oligonuclotide primers given in Table II.

5' tails of forward primers and 3' tails of reverse primers included *att*B1 and *att*B2 sites respectively:

Forward primers: 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTCT-ORF in frame-3' (the TCT sequence preceding the ORF was omitted when the ORF's first coding triplet began with T).

Reverse primers: 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTT-ORF reverse complement-3'.

The number of nucleotides which hybridized to the sequence to be amplified depended on the melting temperature of the primers, which was determined as described by Breslauer *et al.* [*PNAS USA* (1986) 83:3746-50]. The average melting temperature of the selected oligos was 50-55°C for the hybridizing region and 80-85°C for the whole oligos.

C) Amplification

The standard PCR protocol was as follows: 50 ng genomic DNA were used as template in the presence of 0.5 μM each primer, 200 μM each dNTP, 1.5 mM $MgCl_2$, 1x buffer minus Mg^{++} (Gibco-BRL) and 2 units of Taq DNA polymerase (Platinum Taq, Gibco-BRL) in a final volume of 100 μl. Each sample underwent a double-step of amplification: 5 cycles performed using as the hybridizing temperature 50°C, followed by 25 cycles at 68°C.

The standard cycles were as follows:

	Denaturation: 94°C, 2 min
5 cycles:	Denaturation: 94°C, 30 seconds
	Hybridization: 50°C, 50 seconds
	Elongation: 72°C, 1 min. or 2 min. and 40 sec.
25 cycles :	Denaturation: 94°C, 30 seconds
	Hybridization: 68°C, 50 seconds

Elongation: 72°C, 1 min. or 2 min. and 40 sec.

Elongation time was 1 minute for ORFs shorter than 2000bp and 2:40 minutes for ORFs longer than 2000bp. Amplifications were performed using a Gene Amp PCR system 9600 (Perkin Elmer).

To check amplification results, 2µl of each PCR product were loaded onto 1-1.5 agarose gel and the size of amplified fragments was compared with DNA molecular weight standards (DNA marker IX Roche, 1kb DNA ladder Biolabs).

Single band PCR products were purified by PEG precipitation: 300 µl of TE buffer and 200 µl of 30% PEG 8000/30 mM $MgCl_2$ were added to 100 µl PCR reaction. After vortexing, the DNA was centrifuged for 20 min at 10000*g*, washed with 1 vol. 70% ethanol and the pellet dissolved in 30 µl TE. PCR products smaller than 350 bp were purified using a PCR purification Kit (Qiagen) and eluted with 30 µl of the provided elution buffer.

In order to evaluate the yield, 2µl of the purified DNA were subjected to agarose gel electrophoresis and compared to titrated molecular weight standards.

D) Cloning of PCR products into expression vectors

Cloning was performed following the Gateway™ technology's "one-tube protocol", which consists of a two step reaction (BP and LR) for direct insertion of PCR products into expression vectors.

BP reaction (*att*B x *att*P sites): The reaction allowed insertion of the PCR product into a pDONR vector. The pDONR™ 201 vector we used contains the killer toxin gene *ccdB* between *att*P1 and *att*P2 sites to minimize background colonies lacking the PCR insert, and a selectable marker gene for kanamycin resitance. The reaction resulted in a so called pEntry vector, in which the GBS gene was located between *att*L1 and *att*L2 sites.

60 fmol of PCR product and 100 ng of pDONR™ 201 vector were incubated with 2.5 µl of BP clonase™ in a final volume of 12.5 µl for 4 hours at 25°C.

LR reaction (*att*L x *att*R sites): The reaction allowed the insertion of the GBS gene, now present in the pEntry vector, into *E.coli* expression vectors (pDestination vectors, containing *att*R sites). Two pDestination vectors were used (pDEST15 for N- terminal GST fusions – Figure 86; and pDEST17-1 for N-terminal His-tagged fusions – Figure 87). Both allow transcription of the ORF fusion coding mRNA under T7 RNA polymerase promoter [Studier *et al* (1990) *Meth. Enzymol* 185: 60*ff*].

To 5 µl of BP reaction were added 0.25 µl of 0.75 M NaCl, 100 ng of destination vector and 1.5 µl of LR clonase™ . The reaction was incubated at 25°C for 2 hours and stopped with 1 µl of 1 mg/ml proteinase K solution at 37°C for 15 min.

1 µl of the completed reaction was used to transform 50 µl electrocompetent BL21-SI™ cells (0.1 cm, 200 ohms, 25 µF). BL21-SI cells contain an integrated T7 RNA polymerase gene under the control of the salt-inducible *prU* promoter [Gowrishankar (1985) *J. Bacteriol.* 164:434*ff*]. After electroporation cells were diluted in 1ml SOC medium (20 g/l bacto-tryptone, 5 g/l yeast extract, 0.58 g/l NaCl, 0.186 g/l KCl, 20 mM glucose, 10 mM $MgCl_2$) and incubated at 37°C for 1 hour. 200 µl cells were plated onto LBON plates (Luria Broth medium without NaCl) containing 100 µg/ ml ampicillin. Plates were then incubated for 16 hours at 37°C.

Entry clones: In order to allow the future preparation of Gateway compatible pEntry plasmids containing genes which might turn out of interest after immunological assays, 2.5 µl of BP reaction were incubated for 15 min in the presence of 3 µl 0.15 mg/ml proteinase K solution and then kept at –20°C. The reaction was in this way available to transform *E.coli* competent cells so as to produce Entry clones for future introduction of the genes in other Destination vectors.

E) Protein expression

Single colonies derived from the transformation of LR reactions were inoculated as small-scale cultures in 3 ml LBON 100 µg/ml ampicillin for overnight growth at 25°C. 50-200 µl of the culture was inoculated in 3 ml LBON/Amp to an initial OD600 of 0.1. The cultures were grown at 37°C until OD600 0.4-0.6 and recombinant protein expression was induced by adding NaCl to a final concentration of 0.3 M. After 2 hour incubation the final OD was checked and the cultures were cooled on ice. 0.5 OD_{600} of cells were harvested by centrifugation. The cell pellet was suspended in 50 µl of protein Loading Sample Buffer (50 mM TRIS-HCl pH 6.8, 0.5% w/v SDS, 2.5% v/v glycerin, 0.05% w/v Bromophenol Blue, 100 mM DTT) and incubated at 100 °C for 5 min. 10 µl of sample was analyzed by SDS-PAGE and Coomassie Blue staining to verify the presence of induced protein band.

F) Purification of the recombinant proteins

Single colonies were inoculated in 25 ml LBON 100 µg/ml ampicillin and grown at 25°C overnight. The overnight culture was inoculated in 500 ml LBON/amp and grown under shaking at 25 °C until OD_{600} values of 0.4-0.6. Protein expression was then induced by adding NaCl to a final concentration of 0.3 M. After 3 hours incubation at 25 °C the final OD_{600} was checked and the cultures were cooled on ice. After centrifugation at 6000 rpm (JA10 rotor, Beckman) for 20 min., the cell pellet was processed for purification or frozen at –20 °C.

Proteins were purified in 1 of 3 ways depending on the fusion partner and the protein's solubility:

Purification of soluble His-tagged proteins from *E.coli*

1. Transfer pellets from –20°C to ice bath and reconstitute each pellet with 10 ml B-PER™ solution (Bacterial-Protein Extraction Reagent, Pierce cat. 78266), 10 µl of a 100 mM $MgCl_2$ solution, 50

µl of DNAse I (Sigma D-4263, 100 Kunits in PBS) and 100 µl of 100 mg/ml lysozyme in PBS (Sigma L-7651, final concentration 1 mg/ml).

2. Transfer resuspended pellets in 50 ml centrifuge tubes and leave at room temperature for 30-40 minutes, vortexing 3-4 times.
3. Centrifuge 15-20 minutes at about 30-40000 x *g*.
4. Prepare Poly-Prep (Bio-Rad) columns containing 1 ml of Fast Flow Ni-activated Chelating Sepharose (Pharmacia). Equilibrate with 50 mM phosphate buffer, 300 mM NaCl, pH 8.0.
5. Store the pellet at –20°C, and load the supernatant on to the columns.
6. Discard the flow through.
7. Wash with 10 ml 20 mM imidazole buffer, 50 mM phosphate, 300 mM NaCl, pH 8.0.
8. Elute the proteins bound to the columns with 4.5 ml (1.5 ml + 1.5 ml + 1.5 ml) 250 mM imidazole buffer, 50 mM phosphate, 300 mM NaCl, pH 8.0 and collect three fractions of ~1.5 ml each. Add to each tube 15 µl DTT 200 mM (final concentration 2 mM).
9. Measure the protein concentration of the collected fractions with the Bradford method and analyse the proteins by SDS-PAGE.
10. Store the collected fractions at +4°C while waiting for the results of the SDS-PAGE analysis.
11. For immunisation prepare 4-5 aliquots of 20-100 µg each in 0.5 ml in 40% glycerol. The dilution buffer is the above elution buffer, plus 2 mM DTT. Store the aliquots at –20°C until immunisation.

Purification of His-tagged proteins from inclusion bodies

1. Bacteria are collected from 500 ml cultures by centrifugation. If required store bacterial pellets at -20°C. Transfer the pellets from -20°C to room temperature and reconstitute each pellet with 10 ml B-PER™ solution, 10 µl of a 100 mM $MgCl_2$ solution (final 1 mM), 50 µl of DNAse I equivalent to 100 Kunits units in PBS and 100 µl of a 100 mg/ml lysozime (Sigma L-7651) solution in PBS (equivalent to 10 mg, final concentration 1 mg/ml).
2. Transfer the resuspended pellets in 50 ml centrifuge tubes and let at room temperature for 30-40 minutes, vortexing 3-4 times.
3. Centrifuge 15 minutes at 30-4000 x g and collect the pellets.
4. Dissolve the pellets with 50 mM TRIS-HCl, 1 mM TCEP {Tris(2-carboxyethyl)-phosphine hydrochloride, Pierce} , 6M guanidine hydrochloride, pH 8.5. Stir for ~ 10 min. with a magnetic bar.
5. Centrifuge as described above, and collect the supernatant.
6. Prepare Poly-Prep (Bio-Rad) columns containing 1 ml of Fast Flow Ni-activated Chelating Sepharose (Pharmacia). Wash the columns twice with 5 ml of H_20 and equilibrate with 50 mM TRIS-HCl, 1 mM TCEP, 6M guanidine hydrochloride, pH 8.5.

7. Load the supernatants from step 5 onto the columns, and wash with 5 ml of 50 mM TRIS-HCl buffer, 1 mM TCEP, 6M urea, pH 8.5
8. Wash the columns with 10 ml of 20 mM imidazole, 50 mM TRIS-HCl , 6M urea, 1 mM TCEP, pH 8.5. Collect and set aside the first 5 ml for possible further controls.
9. Elute proteins bound to columns with 4.5ml buffer containing 250 mM imidazole, 50 mM TRIS-HCl, 6M urea, 1 mM TCEP, pH 8.5. Add the elution buffer in three 1.5 ml aliquots, and collect the corresponding three fractions. Add to each fraction 15 µl DTT (final concentration 2 mM).
10. Measure eluted protein concentration with Bradford method and analyse proteins by SDS-PAGE.
11. Dialyse overnight the selected fraction against 50 mM Na phosphate buffer, pH 8.8, containing 10% glycerol, 0.5 M arginine, 5 mM reduced glutathione, 0.5 mM oxidized glutathione, 2 M urea.
12. Dialyse against 50 mM Na phosphate buffer, pH 8.8, containing 10% glycerol, 0.5 M arginine, 5 mM reduced glutathione, 0.5 mM oxidized glutathione.
13. Clarify the dialysed protein preparation by centrifugation and discard the non-soluble material and measure the protein concentration with the Bradford method.
14. For each protein destined to the immunization prepare 4-5 aliquot of 20-100 µg each in 0.5 ml after having adjusted the glycerol content up to 40%. Store the prepared aliquots at –20° C until immunization.

Purification of GST-fusion proteins from *E.coli*

1. Bacteria are collected from 500 ml cultures by centrifugation. If required store bacterial pellets at -20°C. Transfer the pellets from –20°C to room temperature and reconstitute each pellet with 10 ml B-PER™ solution, 10 µl of a 100 mM $MgCl_2$ solution (final 1 mM), 50 µl of DNAse I equivalent to 100 Kunits units in PBS and 100 µl of a 100 mg/ml lysozime (Sigma L-7651) solution in PBS (equivalent to 10 mg, final concentration 1 mg/ml).
2. Transfer the resuspended pellets in 50 ml centrifuge tubes and let at room temperature for 30-40 minutes, vortexing 3-4 times.
3. Centrifuge 15-20 minutes at about 30-40000 x *g*.
4. Discard centrifugation pellets and load supernatants onto the chromatography columns, as follows.
5. Prepare Poly-Prep (Bio-Rad) columns containing 0.5 ml of Glutathione-Sepharose 4B resin. Wash the columns twice with 1 ml of H_20 and equilibrate with 10 ml PBS, pH 7.4.
6. Load supernatants on to the columns and discard the flow through.
7. Wash the columns with 10 ml PBS, pH 7.4.
8. Elute proteins bound to columns with 4.5 ml of 50 mM TRIS buffer, 10 mM reduced glutathione, pH 8.0, adding 1.5 ml + 1.5 ml + 1.5 ml and collecting the respective 3 fractions of ~1.5 ml each.

9. Measure protein concentration of the fractions with the Bradford method and analyse the proteins by SDS-PAGE.
10. Store the collected fractions at +4°C while waiting for the results of the SDS-PAGE analysis.
11. For each protein destined for immunisation prepare 4-5 aliquots of 20-100 µg each in 0.5 ml of 40% glycerol. The dilution buffer is 50 mM TRIS-HCl, 2 mM DTT, pH 8.0. Store the aliquots at –20°C until immunisation.

Figures 167 to 170 and 238 to 239

For the experiments shown in Figures 167 to 170, Figure 238 and lanes 2-6 of Figure 239, the GBS proteins were fused at the N-terminus to thioredoxin and at C-terminus to a poly-His tail. The plasmid used for cloning is pBAD-DEST49 (Invitrogen Gateway™ technology) and expression is under the control of an L(+)-Arabinose dependent promoter. For the production of these GBS antigens, bacteria are grown on RM medium (6g/l Na_2HPO_4, 3g/l KH_2PO_4, 0.5 g/l NaCl, 1 g/l NH_4Cl, pH7,4, 2% casaminoacids, 0.2 % glucose, 1 mM $MgCl_2$) containing 100 µg/ml ampicillin. After incubation at 37°C until cells reach OD_{600}=0.5, protein expression is induced by adding 0.2% (v/v) L(+)Arabinose for 3 hours.

Immunisations with GBS proteins

The purified proteins were used to immunise groups of four CD-1 mice intraperitoneally. 20 µg of each purified protein was injected in Freund's adjuvant at days 1, 21 & 35. Immune responses were monitored by using samples taken on day 0 & 49. Sera were analysed as pools of sera from each group of mice.

FACScan bacteria Binding Assay procedure.

GBS serotype V 2603 V/R strain was plated on TSA blood agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the plates using a sterile dracon swab and inoculated into 100ml Todd Hewitt Broth. Bacterial growth was monitored every 30 minutes by following OD_{600}. Bacteria were grown until OD_{600} = 0.7-0.8. The culture was centrifuged for 20 minutes at 5000rpm. The supernatant was discarded and bacteria were washed once with PBS, resuspended in ½ culture volume of PBS containing 0.05% paraformaldehyde, and incubated for 1 hour at 37°C and then overnight at 4°C.

50µl bacterial cells (OD_{600} 0.1) were washed once with PBS and resuspended in 20µl blocking serum (Newborn Calf Serum, Sigma) and incubated for 20 minutes at room temperature. The cells were then incubated with 100µl diluted sera (1:200) in dilution buffer (20% Newborn Calf Serum 0.1% BSA in PBS) for 1 hour at 4°C. Cells were centrifuged at 5000rpm, the supernatant aspirated and cells washed by adding 200µl washing buffer (0.1% BSA in PBS). 50µl R-Phicoerytrin conjugated $F(ab)_2$ goat anti-mouse, diluted 1:100 in dilution buffer, was added to each sample and incubated for 1 hour at 4°C. Cells were spun down by centrifugation at 5000rpm and washed by adding 200µl of washing buffer. The

supernatant was aspirated and cells resuspended in 200µl PBS. Samples were transferred to FACScan tubes and read. The condition for FACScan setting were: FL2 on; FSC-H threshold:54; FSC PMT Voltage: E 02; SSC PMT: 516; Amp. Gains 2.63; FL-2 PMT: 728. Compensation values: 0.

Samples were considered as positive if they had a Δ mean values > 50 channel values.

Whole Extracts preparation

GBS serotype III COH1 strain and serotype V 2603 V/R strain cells were grown overnight in Todd Hewitt Broth. 1ml of the culture was inoculated into 100ml Todd Hewitt Broth. Bacterial growth was monitored every 30 minutes by following OD_{600}. The bacteria were grown until the OD reached 0.7-0.8. The culture was centrifuged for 20 minutes at 5000 rpm. The supernatant was discarded and bacteria were washed once with PBS, resuspended in 2ml 50mM Tris-HCl, pH 6.8 adding 400 units of Mutanolysin (Sigma-Aldrich) and incubated 3 hrs at 37°C. After 3 cycles of freeze/thaw, cellular debris were removed by centrifugation at 14000*g* for 15 minutes and the protein concentration of the supernatant was measured by the Bio-Rad Protein assay, using BSA as a standard.

Western blotting

Purified proteins (50ng) and total cell extracts (25µg) derived from GBS serotype III COH1 strain and serotype V 2603 V/R strain were loaded on 12% or 15% SDS-PAGE and transferred to a nitrocellulose membrane. The transfer was performed for 1 hours at 100V at 4°C, in transferring buffer (25mM Tris base, 192mM glycine, 20% methanol). The membrane was saturated by overnight incubation at 4°C in saturation buffer (5 % skimmed milk, 0.1% Tween 20 in PBS). The membrane was incubated for 1 hour at room temperature with 1:1000 mouse sera diluted in saturation buffer. The membrane was washed twice with washing buffer (3 % skimmed milk, 0.1% Tween 20 in PBS) and incubated for 1 hour with a 1:5000 dilution of horseradish peroxidase labelled anti-mouse Ig (Bio-Rad). The membrane was washed twice with 0.1% Tween 20 in PBS and developed with the Opti-4CN Substrate Kit (Bio-Rad). The reaction was stopped by adding water.

Unless otherwise indicated, lanes 1, 2 and 3 of blots in the drawings are: (1) the purified protein; (2) GBS-III extracts; and (3) GBS-V extracts. Molecular weight markers are also shown.

In vivo *passive protection assay in neonatal sepsis mouse model.*

The immune sera collected from the CD1 immunized mice were tested in a mouse neonatal sepsis model to verify their protective efficacy in mice challenged with GBS serotype III. Newborn Balb/C littermates were randomly divided in two groups within 24 hrs from birth and injected subcutaneously with 25µl of diluted sera (1:15) from immunized CD1 adult mice. One group received preimmune sera, the other received immune sera. Four hours later all pups were challenged with a 75% lethal dose of the GBS serotype III COH1 strain. The challenge dose obtained diluting a mid log phase culture was administered subcutaneously in 25 µl of saline. The number of pups surviving GBS infection was assessed every 12 hours for 4 days. Results are in Table III.

Example 1

A DNA sequence (GBSx1402) was identified in *S.agalactiae* <SEQ ID 1> which encodes the amino acid sequence <SEQ ID 2>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.48   Transmembrane  169 - 185 ( 169 - 185)

----- Final Results -----
               bacterial membrane --- Certainty=0.1192(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB88235 GB:AL353012 hypothetical serine-rich repeat protein
           [Schizosaccharomyces pombe]
 Identities = 41/152 (26%), Positives = 75/152 (48%), Gaps = 4/152 (2%)

Query: 22  SSIGYADTSDKNTDTSVVTTTLSEEKRSDELDQSSTGSSSENESSSSSEPETNPSTNPPT 81
           SS   +++S +++D+S  ++    E  S+  D SS+ SSSE+ESSS     ++ S++  +
Sbjct: 132 SSDSESESSSEDSDSSSSSSDSESESSSEGSDSSSSSSSSESESSSEDNDSSSSSSDSES 191

Query: 82  TEPSQPSPSEENKPDGRTKTE---IGNNKDISSGTKVLISEDSIKNFSKASSDQEEVDRD 138
              S+ S S  +  D  +++      ++   SS      SED+  + S + S+ E    D
Sbjct: 192 ESSSEDSDSSSSSSDSESESSSEGSDSSSSSSSESESSSEDNDSSSSSSDSESESSSED 251

Query: 139 ESSSSKANDGK-KGHSKPKKELPKTGDSHSDT 169
             SSS ++D + +  SK       + DS  D+
Sbjct: 252 SDSSSSSSDSESESSSKDSDSSSNSSDSEDDS 283
```

There is also homology to SEQ ID 1984.

A related GBS gene <SEQ ID 8785> and protein <SEQ ID 8786> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 5
McG: Discrim Score:      6.72
GvH: Signal Score (-7.5): -4.34
     Possible site: 27
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 1 value:  -0.48 threshold:  0.0
   INTEGRAL    Likelihood = -0.48   Transmembrane  169 - 185 ( 169 - 185)
   PERIPHERAL  Likelihood =  0.16       7
 modified ALOM score:   0.60

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.1192(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif: 159-163
```

SEQ ID 2 (GBS4) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 9 (lane 3; MW 43.1kDa) and Figure 63 (lane 4; MW 50kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 12 (lane 7; MW 30kDa), Figure 63 (lane 3; MW 30kDa) and in Figure 178 (lane 3; MW 30kDa).

GBS4-GST was purified as shown in Figure 190 (lane 6) and Figure 209 (lane 8).

Purified GBS4-His is shown in Figures 89A, 191 (lane 10), 209 (lane 7) and 228 (lanes 9 & 10).

The purified GBS4-His fusion product was used to immunise mice (lane 2 product; 20ìg/mouse). The resulting antiserum was used for Western blot (Figure 89B), FACS, and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2

A DNA sequence (GBSx1100) was identified in *S.agalactiae* <SEQ ID 3> which encodes the amino acid sequence <SEQ ID 4>. This protein is predicted to be aggregation promoting protein. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA69725 GB:Y08498 aggregation promoting protein [Lactobacillus gasseri]
 Identities = 56/103 (54%), Positives = 69/103 (66%), Gaps = 5/103 (4%)

Query: 82  TASQAEAKSQPT-----IENSMNSSSNLSSSDSAAKEEIARRESNGSYTAQNGQYYGRYQ 136
           T S A A+ Q T      + + + + N S S++AAK  +A RES G Y+A NGQY G+YQ
Sbjct: 195 TYSYASAQKQTTQVAQKTQTTTSYTLNASGSEAAAKAWMAGRESGGPYSAGNGQYIGKYQ 254

Query: 137 LSQSYLNGDLSPENQEKVADNYVVSRYGSWSAALSFWNSNGWY 179
           LS SYL GD S  NQE+VADNYV SRYGSW+ A  FW +NGWY
Sbjct: 255 LSASYLGGDYSAANQERVADNYVKSRYGSWTGAQKFWQTNGWY 297
```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8709> and protein <SEQ ID 8710> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 9
McG: Discrim Score:      2.59
GvH: Signal Score (-7.5): -0.42
     Possible site: 33
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:   6.79 threshold:  0.0
   PERIPHERAL  Likelihood =  6.79      59
 modified ALOM score:  -1.86

*** Reasoning Step: 3

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
57.5/71.3% over 92aa
  Lactobacillus gasseri
```

```
  EGAD|154417| aggregation promoting protein Insert characterized
  GP|1619598|emb|CAA69725.1||Y08498 aggregation promoting protein Insert characterized

ORF01056(547 - 837 of 1137)
EGAD|154417|164788(205 - 297 of 297) aggregation promoting protein {Lactobacillus
gasseri}GP|1619598|emb|CAA69725.1||Y08498 aggregat
ion promoting protein {Lactobacillus gasseri}
%Match = 14.6
%Identity = 57.4  %Similarity = 71.3
Matches = 54  Mismatches = 26  Conservative Sub.s = 13

507       537       567       597       627       657       687       717
SLNSISNADVISIGDVLKLDNSTASQAEAKSQPTIENSMNSSSNLSSSDSAAKEEIARRESNGSYTAQNGQYYGRYQLSQ
::    :| |              :|   |:| |   ::|:|     |::|||  :| ||| | |:| |||| |:||||
NVQRTYSAPVQQRTYSYASAQKQTTQVAQKTQTTTSYTLNASG----SEAAAKAWMAGRESGGPYSAGNGQYIGKYQLSA
                 200       210       220           230       240       250

747       777       807       837       867       897       927       957
SYLNGDLSPENQEKVADNYVVSRYGSWSAALSFWNSNGWY**KLIKQRDLLKIKSLCNIFNIYSIAR*QIKYNIGNMNKR
||| || |  |||:|||||| ||||||: |  || :||||
SYLGGDYSAANQERVADNYVKSRYGSWTGAQKFWQTNGWY
          270       280       290
```

A related GBS gene <SEQ ID 8711> and protein <SEQ ID 8712> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 9
McG: Discrim Score:      2.59
GvH: Signal Score (-7.5): -0.42
     Possible site: 33
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:   6.79 threshold:  0.0
   PERIPHERAL  Likelihood =  6.79      59
 modified ALOM score:  -1.86

*** Reasoning Step: 3

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
44.0/62.0% over 115aa
                                                                  Bacillus subtilis
  EGAD|108478|  hypothetical protein Insert characterized OMNI|NT01BS1100 p60-related
protein Insert characterized
  GP|2226145|emb|CAA74437.1||Y14079 hypothetical protein Insert characterized
  GP|2633272|emb|CAB12776.1||Z99109  similar  to  cell  wall-binding  protein  Insert
characterized
  PIR|B69825|B69825 cell wall-binding protein homolog yhdD - Insert characterized

ORF01746(340 - 633 of 954)
EGAD|108478|BS0936(57 - 172 of 488) hypothetical protein {Bacillus subtilis}OMNI|NT01BS1100
p60-related  proteinGP|2226145|emb|CAA74437.1||Y14079  hypothetical  protein  {Bacillus
subtilis}GP|2633272|emb|CAB12776.1||Z99109 similar to cell wall-binding protein {Bacillus
subtilis}PIR|B69825|B69825 cell wall-binding protein homolog yhdD - Bacillus subtilis
%Match = 9.0
%Identity = 44.0  %Similarity = 62.0
Matches = 44  Mismatches = 35  Conservative Sub.s = 18

120       150       180       210       240       270       300       330
*DQFMVLAFSFI*CEKLNNFT*RKLKIVFWRPFLY*FTIYL**ISSKAKQLVIFTRYDSTRIN**KRAYIMSITSVKKSK

                                MKKKLAAGLTASAIVGTTLVVTPAEAATIKVKSGDSLWKLAQTYNTSVAALTS
                                        10        20        30        40        50
```

```
360       390                              435       465       495       525
PFKLGVAGLLVGASLALPLSVSAAS---------------YTVKSGDTLSAIAKNHKTTVQELVSLNSISNADVISIGDV
   |    | :| :| :| | |: |                |||||||:|  ||   | |||||  || :| :|:|  |
ANHLSTTVLSIGQTLTIPGSKSSTSSSTSSSTTMKSGSSVYTVKSGDSLWLIANEFKMTVQELKKLNGLS-SDLIRAGQK
               70        80        90       100       110       120       130

543       573       603       633       663       693       723       753
LKLD----NSTASQAEAKSQPTIENSMNSSSNLSSSDSAAKEEIASS*IKXVVILHRMDNIMEDINCLNLT*MATYLLKI
||:     :|::|  :: |  :  :|  |||| ||| |::           :  :    |:      :     :  :
LKVSGTVSSSSSSSKKSNSNKSSSSSSKSSSNKSSSSSSSTGTYKVQLGDSLWKIANKVNMSIAELKVLNNLKSDTIYVN
               150       160       170       180       190       200       210
```

SEQ ID 8712 (GBS166) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 30 (lane 2; MW 13.1kDa).

The GBS166-His fusion product was purified (Figure 200, lane 10) and used to immunise mice. The resulting antiserum was used for FACS (Figure 315), which confirmed that the protein is immunoaccessible on GBS bacteria.

SEQ ID 4 (GBS15) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 9 (lane 5; MW 44.8kDa), Figure 63 (lane 5; MW 44.8kDa) and Figure 66 (lane 7; MW 45kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 10 (lane 4; MW 22.3kDa). It was also expressed as GBS15L, with SDS-PAGE analysis of total cell extract is shown in Figure 185 (lane 1; MW 50kDa).

Purified GBS15-GST is shown in Figure 91A, Figure 190 (lane 9), Figure 210 (lane 4) and Figure 245 (lanes 4 & 5).

The purified GBS15-GST fusion product was used to immunise mice (lane 1 + 2 products; 20μg/mouse). The resulting antiserum was used for Western blot (Figure 91B), FACS (Figure 91C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 3

A DNA sequence (GBSx0091) was identified in *S.agalactiae* <SEQ ID 303> which encodes the amino acid sequence <SEQ ID 304>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.66   Transmembrane   22 -  38 (  15 -  41)

----- Final Results -----
               bacterial membrane --- Certainty=0.4864(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA72096 GB:Y11213 hypothetical protein [Streptococcus thermophilus]
 Identities = 149/274 (54%), Positives = 208/274 (75%), Gaps = 9/274 (3%)

Query: 23  FLVSLLLSFGIFSLIIPKSNP--KLTKKDFLTKKVIPLNYVALGDSLTEGVGDTTSQGGF 80
```

```
           F +  LL  GI   IIP S+   K++ K    KK   + YVA+GDSLT+GVGD+++QGGF
Sbjct: 5   FFLLFLLFVGILIFIIPSSHQSSKISDKIRSVKKE-KVTYVAIGDSLTQGVGDSSNQGGF 63

Query: 81  VPLLSESLHNRYSYQVTSVNYGVSGNTSQQILKRMTTDPQIEKDLEKADLLTLTVGGNDV 140
           VP+LS++L + +++QVT  NYG++GNTS QILKRM     I++DL+KA L+TLTVGGNDV
Sbjct: 64  VPVLSQALESDFNWQVTPRNYGIAGNTSNQILKRMQEKKDIKRDLKKAKLMTLTVGGNDV 123

Query: 141 LAVIRKELSHLSLNSFEKPAEAYKERLKEILAKARQDNPKLPIYVLGIYNPFYLNFPQLT 200
           + VI+  +++L++N+F K A  Y++RL++I+  AR++N  LPIY++GIYNPFYLNFP++T
Sbjct: 124 IHVIKDNITNLNVNTFSKAAVDYQKRLRQIIELARKENKTLPIYIIGIYNPFYLNFPEMT 183

Query: 201 KMQTVIDNWNKATKEVVDASENVYFVPINDRLYKGINGKEGITES------SNSQASITN 254
           +MQT++DNWN++T+EV    +NVYFVP+ND LYKGINGK G+T S      + S     N
Sbjct: 184 EMQTIVDNWNRSTEEVSKEYDNVYFVPVNDLLYKGINGKGGVTSSDETSQPTKSSQDSLN 243

Query: 255 DALFTGDHFHPNNIGYQIMSNAVMEKINETRKNW 288
           DALF  DHFHPNN GYQIMS+A++++IN+T+K W
Sbjct: 244 DALFEEDHFHPNNTGYQIMSDAILKRINQTKKEW 277
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 305> which encodes the amino acid sequence <SEQ ID 306>. Analysis of this protein sequence reveals the following:

```
Possible site: 39

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.05   Transmembrane   18 -  34 (  10 -  37)

----- Final Results -----
               bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9123> which encodes the amino acid sequence <SEQ ID 9124>. Analysis of this protein sequence reveals the following:

```
     Possible site: 33
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.05   Transmembrane   12 -  28

----- Final Results -----
               bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 178/282 (63%), Positives = 218/282 (77%)

Query: 5   LLLWFVMNKKKILTGLSFFLVSLLLSFGIFSLIIPKSNPKLTKKDFLTKKVIPLNYVALG 64
           L LWFVMN + + +G+ FF++SL L+F + ++IIPKSN +L K DFL K+ + + YVA+G
Sbjct: 1   LRLWFVMNNRHLFSGIFFFVISLCLAFLLLNIIIPKSNSRLKKSDFLKKEQVAIQYVAIG 60

Query: 65  DSLTEGVGDTTSQGGFVPLLSESLHNRYSYQVTSVNYGVSGNTSQQILKRMTTDPQIEKD 124
           DSLTEGVGD T QGGFVPLL+  L   +   V   NYGVSG+TSQQIL RM    QI+
Sbjct: 61  DSLTEGVGDLTHQGGFVPLLTNDLSEYFKANVNHQNYGVSGDTSQQILDRMIKQKQIQLS 120

Query: 125 LEKADLLTLTVGGNDVLAVIRKELSHLSLNSFEKPAEAYKERLKEILAKARQDNPKLPIY 184
           L+KAD++TLTVGGNDV+AVIRK L+ L ++SF KPA  Y++RL++I+  AR+DN  LPI+
Sbjct: 121 LKKADIMTLTVGGNDVMAVIRKNLADLQVSSFRKPARQYQKRLRQIIELARKDNKDLPIF 180

Query: 185 VLGIYNPFYLNFPQLTKMQTVIDNWNKATKEVVDASENVYFVPINDRLYKGINGKEGITE 244
           +LGIYNPFYLNFP+LT MQ VID+WN  TKEVV   + VYFVPIND LYKGING+EGI
Sbjct: 181 ILGIYNPFYLNFPELTDMQKVIDDWNTKTKEVVGEYDRVYFVPINDLLYKGINGQEGIVH 240

Query: 245 SSNSQASITNDALFTGDHFHPNNIGYQIMSNAVMEKINETRK 286
           SS  Q +I NDALFTGDHFHPNN GYQIMSNAVMEKI +  K
```

```
Sbjct: 241 SSGDQTTIVNDALFTGDHFHPNNTGYQIMSNAVMEKIKKHEK 282
```

A related GBS gene <SEQ ID 5> and protein <SEQ ID 6> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 4
SRCFLG: 0
McG: Length of UR:   24
     Peak Value of UR:   3.02
     Net Charge of CR: 3
McG: Discrim Score:     12.27
GvH: Signal Score (-7.5): -3.44
     Possible site: 22
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program   count: 1 value:  -9.66 threshold:  0.0
   INTEGRAL    Likelihood = -9.66   Transmembrane   12 -  28 (   5 -  31)
   PERIPHERAL  Likelihood =  1.96     118
 modified ALOM score:   2.43
icm1 HYPID: 7  CFP: 0.486

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4864(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
56.0/80.3% over 272aa
GP|1850894| hypothetical protein Insert characterized

ORF02006(367 - 1164 of 1467)
GP|1850894|emb|CAA72096.1||Y11213(5 - 277 of 280) hypothetical protein {Streptococcus
thermophilus}
%Match = 30.8
%Identity = 56.0  %Similarity = 80.2
Matches = 150  Mismatches = 49  Conservative Sub.s = 65

141       171       201       231       261       291       321       351
AV*RPSANG*IILLKVPKHEKLLKLASPTVVKLIWLITLEKN*LF*VLLYPF*KLAQSSKLILVRMHLLLWFVMNKKKIL

381       411       435       465       495       525       555       585
TGLSFFLVSLLLSFGIFSLIIPKSN--PKLTKKDFLTKKVIPLNYVALGDSLTEGVGDTTSQGGFVPLLSESLHNRYSYQ
  :: |:: :||  ||: :||| |:   |:: |    ||   : |||:|||||:||||:::||||||:||::| : :::|
 SFAGFFLLFLLFVGILIFIIPSSHQSSKISDKIRSVKK-EKVTYVAIGDSLTQGVGDSSNQGGFVPVLSQALESDFNWQ
         10        20        30        40        50        60        70

615       645       675       705       735       765       795       825
VTSVNYGVSGNTSQQILKRMTTDPQIEKDLEKADLLTLTVGGNDVLAVIRKELSHLSLNSFEKPAEAYKERLKEILAKAR
||  |||::|||| ||||||     |::||:|| |:|||||||||: ||:  :::|::|:| | |  |::||::|:  ||
VTPRNYGIAGNTSNQILKRMQEKKDIKRDLKKAKLMTLTVGGNDVIHVIKDNITNLNVNTFSKAAVDYQKRLRQIIELAR
          90        100       110       120       130       140       150

855       885       915       945       975      1005                1044
QDNPKLPIYVLGIYNPFYLNFPQLTKMQTVIDNWNKATKEVVDASENVYFVPINDRLYKGINGKEGIT-------ESSNS
::|  ||||::|||||||||||::|:|||::||||::|:||    :||||||:|| ||||||||  |:|        : :|
KENKTLPIYIIGIYNPFYLNFPEMTEMQTIVDNWNRSTEEVSKEYDNVYFVPVNDLLYKGINGKGGVTSSDETSQPTKSS
         170       180       190       200       210       220       230

1074      1104      1134      1164      1194      1224      1254      1284
QASITNDALFTGDHFHPNNIGYQIMSNAVMEKINETRKNWP*FKFLEMGISLIVGN*PFLHSSDCKSLNSST*A*YRKNF
| |: |||||  |||||||  ||||||:|::::||:|:| |
QDSL-NDALFEEDHFHPNNTGYQIMSDAILKRINQTKKEWSGE
          250       260       270       280
```

SEQ ID 6 (GBS103) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 36 (lane 4; MW 32kDa).

The GBS103-His fusion product was purified (Figure 107A; see also Figure 201, lane 9) and used to immunise mice (lane 2+3 product; 18.5µg/mouse). The resulting antiserum was used for Western blot (Figure 107B), FACS (Figure 107C) and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 4

A DNA sequence (GBSx1316) was identified in *S.agalactiae* <SEQ ID 3837> which encodes the amino acid sequence <SEQ ID 3838>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.30   Transmembrane 1058 -1074 (1056 -1075)

----- Final Results -----
               bacterial membrane --- Certainty=0.2720(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 7> and protein <SEQ ID 8> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:    -13.26
GvH: Signal Score (-7.5): -5.76
     Possible site: 41
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -4.30 threshold:  0.0
   INTEGRAL    Likelihood = -4.30   Transmembrane  489 - 505 ( 487 - 506)
   PERIPHERAL  Likelihood =  3.71       97
 modified ALOM score:   1.36

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.2720(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif: 478-482
```

SEQ ID 8 (GBS195) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 24 (lane 8). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 31 (lane 5).

GBS195C was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 175 (lane 6 & 7; MW 81kDa).

GBS195L was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 83 (lane 2; MW 123kDa).

GBS195LN was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 83 (lane 3; MW 66kDa).

GBS195-GST was purified as shown in Figure 198, lane 5. GBS195-His was purified as shown in Figure 222, lane 4-5. GBS195N-His was purified as shown in Figure 222, lane 6-7.

The GBS195-GST fusion product was purified (Figure 87A) and used to immunise mice (lane 1 product; 13.6μg/mouse). The resulting antiserum was used for Western blot (Figure 87B), FACS, and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 5

A DNA sequence (GBSx0002) was identified in *S.agalactiae* <SEQ ID 4043> which encodes the amino acid sequence <SEQ ID 4044>. This protein is predicted to be lipoprotein MtsA. Analysis of this protein sequence reveals the following:

```
Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3361(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9403> which encodes amino acid sequence <SEQ ID 9404> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3177> which encodes the amino acid sequence <SEQ ID 3178>. Analysis of this protein sequence reveals the following:

```
      Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2412(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 146/168 (86%), Positives = 161/168 (94%)

Query: 1   MNLENGIIYSKNIAKQLIAKDPKNKATYEKNRDAYVAKLEKLDKEAKSKFNAIPANKKLI 60
           +NLENGIIYSKNIAKQLIAKDPKNK TYEKN  AYVAKLEKLDKEAKSKF+AI  NKKLI
Sbjct: 107 LNLENGIIYSKNIAKQLIAKDPKNKETYEKNLKAYVAKLEKLDKEAKSKFDAIAENKKLI 166

Query: 61  VTSEGCFKYFSKAYGVPSAYIWEINTEEEGTPDQITSLVKKLKQVRPSALFVESSVDKRP 120
           VTSEGCFKYFSKAYGVPSAYIWEINTEEEGTPDQI+SL++KLK ++PSALFVESSVD+RP
Sbjct: 167 VTSEGCFKYFSKAYGVPSAYIWEINTEEEGTPDQISSLIEKLKVIKPSALFVESSVDRRP 226
```

```
Query: 121 MKSVSRESGIPIYAEIFTDSIAKKGQKGDSYYAMMKWNLDKIAEGLAK 168
           M++VS++SGIPIY+EIFTDSIAKKG+ GDSYYAMMKWNLDKI+EGLAK
Sbjct: 227 METVSKDSGIPIYSEIFTDSIAKKGKPGDSYYAMMKWNLDKISEGLAK 274
```

SEQ ID 9404 (GBS679) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 164 (lane 7-9; MW 36kDa) and in Figure 188 (lane 8; MW 36kDa). Purified protein is shown in Figure 242, lanes 9 & 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 6

A DNA sequence (GBSx0003) was identified in *S.agalactiae* <SEQ ID 8485> which encodes the amino acid sequence <SEQ ID 8486>. This protein is predicted to be ATP-binding protein MtsB. Analysis of this protein sequence reveals the following:

```
Possible site: 55

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2097(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 8765> which encodes the amino acid sequence <SEQ ID 8766>. Analysis of this protein sequence reveals the following:

```
Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1929(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 143/238 (60%), Positives = 186/238 (78%), Gaps = 2/238 (0%)

Query: 1   MIISKHLSVSYDNNL-VLEDINLRLEGSGIIGILGPNGAGKSTLMKALLGLVDSTGESGI 59
           MI + +L V+YD N   LE IN+ +EG  I+GI+GPNGAGKST MKA+L L+D  G   +
Sbjct: 10  MITTNNLCVTYDGNSNALEAINVTIEGPSIVGIIGPNGAGKSTFMKAILNLIDYQGHVTV 69

Query: 60  GG-DLLPLMGRVAYVEQKTNIDYQFPITVGECVSLGLYKERGLFKRLSKTDWEKVSRVID 118
            G D   L   VAYVEQ++ IDY FPITV ECV+LG Y + GLF+R+ K  +E+V +V+
Sbjct: 70  DGKDGRKLGHTVAYVEQRSMIDYNFPITVKECVALGTYSKLGLFRRVGKKQFEQVDKVLK 129

Query: 119 QVGLRGFENRPINALSGGQFQRMLMARCLVQEADYIFLDEPFVGIDSISEQIIVNLLKKL 178
           QVGL  F +RPI +LSGGQFQRML+ARCL+QE+DYIFLDEPFVGIDS+SE+IIV+LLK+L
Sbjct: 130 QVGLEDFGHRPIKSLSGGQFQRMLVARCLIQESDYIFLDEPFVGIDSVSEKIIVDLLKEL 189

Query: 179 SKAGKLILVVHHDLSKVDHYFDQVIILNRHLIACGPIDQAFTRENLSAAYGDAILLGQ 236
             AGK IL+VHHDLSKV+HYFD+++ILN+HL+A G + + FT + LS AYG+ ++LG+
Sbjct: 190 KMAGKTILIVHHDLSKVEHYFDKLMILNKHLVAYGNVCEVFTVDTLSKAYGNHLILGK 247
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 7

A DNA sequence (GBSx0004) was identified in *S.agalactiae* <SEQ ID 9> which encodes the amino acid sequence <SEQ ID 10>. Analysis of this protein sequence reveals the following:

```
Possible site: 28

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 8

A DNA sequence (GBSx0005) was identified in *S.agalactiae* <SEQ ID 11> which encodes the amino acid sequence <SEQ ID 12>. This protein is predicted to be integral membrane protein MtsC (znuB). Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 6
McG: Discrim Score:      3.77
GvH: Signal Score (-7.5): -0.47
     Possible site: 45
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-10.83   Transmembrane  138 - 154 ( 134 - 162)
   INTEGRAL    Likelihood = -7.96   Transmembrane   60 -  76 (  50 -  86)
   INTEGRAL    Likelihood = -6.95   Transmembrane   95 - 111 (  93 - 118)
   INTEGRAL    Likelihood = -5.79   Transmembrane  180 - 196 ( 174 - 216)
   INTEGRAL    Likelihood = -4.35   Transmembrane  198 - 214 ( 197 - 216)
   INTEGRAL    Likelihood = -4.30   Transmembrane  250 - 266 ( 246 - 268)
   INTEGRAL    Likelihood = -3.93   Transmembrane  222 - 238 ( 221 - 241)
   PERIPHERAL  Likelihood =  5.94     116
   modified ALOM score:   2.67

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5331(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 13> which encodes the amino acid sequence <SEQ ID 14>. Analysis of this protein sequence reveals the following:

```
     Possible site: 45
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-11.25   Transmembrane  138 - 154 ( 134 - 163)
   INTEGRAL    Likelihood = -9.08   Transmembrane   66 -  82 (  50 -  86)
   INTEGRAL    Likelihood = -6.79   Transmembrane   95 - 111 (  93 - 118)
   INTEGRAL    Likelihood = -5.63   Transmembrane  180 - 196 ( 176 - 216)
   INTEGRAL    Likelihood = -4.73   Transmembrane  221 - 237 ( 218 - 241)
   INTEGRAL    Likelihood = -4.35   Transmembrane  250 - 266 ( 246 - 268)
   INTEGRAL    Likelihood = -4.35   Transmembrane  198 - 214 ( 197 - 216)
   INTEGRAL    Likelihood = -2.81   Transmembrane   48 -  64 (  47 -  64)

----- Final Results -----
```

```
                bacterial membrane --- Certainty=0.5501(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 224/275 (81%), Positives = 255/275 (92%)

Query: 1   MFTKFFEGLLTYHFLQNAFITAIVIGIVAGAVGCFIILRSMSLMGDAISHAVLPGVAISF 60
           M  KFFEGL++YHFLQNA ITA+VIGIV+GAVGCFIILRSMSLMGDAISHAVLPGVA+SF
Sbjct: 1   MSMKFFEGLMSYHFLQNALITAVVIGIVSGAVGCFIILRSMSLMGDAISHAVLPGVALSF 60

Query: 61  ILGINFFIGAIVFGLLSSIIITYIKENSVIKGDTAIGITFSSFLALGIILIGLANSTTDL 120
           ILG+NFFIGAI+FGLL+S+IITYIKENSVIKGDTAIGITFSSFLALG+ILIG+ANS+TDL
Sbjct: 61  ILGVNFFIGAIIFGLLASVIITYIKENSVIKGDTAIGITFSSFLALGVILIGVANSSTDL 120

Query: 121 FHILFGNILAVQDSDKYMTIIVGLIVLTLITIFFKELLLTSFDPVLAKSMGMRVSFYHYL 180
           FHILFGNILAVQDSDK++TI V + VL +I++FFKELLLTSFDP+LAKSMG++V+ YHYL
Sbjct: 121 FHILFGNILAVQDSDKWITIGVSIFVLVVISLFFKELLLTSFDPILAKSMGVKVNAYHYL 180

Query: 181 LMILLTLVAVTAMQSVGTILIVALLITPAATAYLYVKSLRTMLFLSSALGAVASVLGLYI 240
           LM+LLTLVAVTAMQSVGTILIVALLITPAATAYLY  SL+ ML +SS LGA+ASVLGLY+
Sbjct: 181 LMVLLTLVAVTAMQSVGTILIVALLITPAATAYLYANSLKVMLVMSSLLGALASVLGLYL 240

Query: 241 GYTFNIAAGSSIVLTSTFMFLLAFLFSPKQSLFKK 275
           GYTFN+AAGSSIVLTS  MFL++F  SPKQ   K+
Sbjct: 241 GYTFNVAAGSSIVLTSAMMFLISFFVSPKQGYLKR 275
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 9

A DNA sequence (GBSx0006) was identified in *S.agalactiae* <SEQ ID 15> which encodes the amino acid sequence <SEQ ID 16>. Analysis of this protein sequence reveals the following:

```
Possible site: 38

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1280(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 10

A DNA sequence (GBSx0007) was identified in *S.agalactiae* <SEQ ID 17> which encodes the amino acid sequence <SEQ ID 18>. This protein is predicted to be peptidyl-prolyl cis-trans isomerase 10 (rotamase). Analysis of this protein sequence reveals the following:

```
Lipop Possible site: 19   Crend: 2
McG: Discrim Score:      5.27
GvH: Signal Score (-7.5): -4.14
     Possible site: 19
>>> May be a lipoprotein
```

```
ALOM program   count: 0 value:   9.34 threshold:  0.0
   PERIPHERAL  Likelihood =  9.34      89
 modified ALOM score:  -2.37

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA19257 GB:AL023704 putative Cyclophilin-type peptidyl-prolyl
          cis-trans isomerase protein [Schizosaccharomyces pombe]
 Identities = 88/224 (39%), Positives = 123/224 (54%), Gaps = 46/224 (20%)

Query: 50  NKKTKQALKADKKAFPQLDKAVAKNEAQ-----------VLIKTSKGDINIKLFPKYAPL 98
           N  TK  L +D+  + +    V  NE +             +I T++GDI+IKL+P+ AP
Sbjct: 419 NMSTKFTL-SDRDVYNEQVLPVTNNEGRQENGNILLGKAAIIHTTQGDISIKLYPEEAPK 477

Query: 99  AVENFLTHAKEGYYNGLSFHRVIKDFMIQSGDPNGDGTGGKSIWNSKDKKKDSGNGFVNE 158
           AV+NF THA+ GYY+   FHR+IK+FMIQ GDP GDGTGG+SIW     KKD    F +E
Sbjct: 478 AVQNFTTHAENGYYDNTIFHRIIKNFMIQGGDPLGDGTGGESIW-----KKD----FEDE 528

Query: 159 ISPYLYNIRG-SLAMANAGADTNGSQFFINQSQQDHSKQLSDKKVPKVIIKAYSEGGNPS 217
           ISP L + R  +++MAN+G +TNGSQFFI                               P
Sbjct: 529 ISPNLKHDRPFTVSMANSGPNTNGSQFFITTDL-----------------------TPW 564

Query: 218 LDGGYTVFGQVISGMETVDKIASVEVTKSDQPKEKITITSIKVI 261
           LDG +T+F +  +G++ V +I   E  K D+P E   I +I ++
Sbjct: 565 LDGKHTIFARAYAGLDVVHRIEQGETDKYDRPLEPTKIINISIV 608
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 19> which encodes the amino acid sequence <SEQ ID 20>. Analysis of this protein sequence reveals the following:

```
     Possible site: 19

>>> May be a lipoprotein

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB88542 GB:AL353818 putative protein [Arabidopsis thaliana]
 Identities = 83/186 (44%), Positives = 104/186 (55%), Gaps = 34/186 (18%)

Query: 78  VVMRTSQGDITLKLFPKYAPLAVENFLTHAKKGYYDNLTFHRVINDFMIQSGDPKGDGTG 137
           V+M T+ GDI +KL+P+  P  VENF TH + GYYDN  FHRVI  FMIQ+GDP GDGTG
Sbjct: 476 VIMHTTLGDIHMKLYPEECPKTVENFTTHCRNGYYDNHLFHRVIRGFMIQTGDPLGDGTG 535

Query: 138 GESIWKGKDPKKDAGNGFVNEISPFLYHIRG-ALAMANAGANTNGSQFYINQNKKNQSKG 196
           G+SIW         G  F +E    L H R   L+MANAG NTNGSQF+I
Sbjct: 536 GQSIW---------GREFEDEFHKSLRHDRPFTLSMANAGPNTNGSQFFITT-------- 578

Query: 197 LSSTNYPKPIISAYEHGGNPSLDGGYTVFGQVIDGMDVVDKIAATSINQNDKPEQDITIT 256
                           P LD  +TVFG+V+ GMDVV  I     ++ND+P QD+ I
Sbjct: 579 ----------------VATPWLDNKHTVFGRVVKGMDVVQGIEKVKTDKNDRPYQDVKIL 622

Query: 257 SIDIVK 262
           ++ + K
Sbjct: 623 NVTVPK 628
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 172/267 (64%), Positives = 221/267 (82%)

Query: 1   MKKIIYLGLACVSILTLSGCESIERSLKGDRYVDQKLAENSSKEATEQLNKKTKQALKAD 60
           MKK++ L L  +S+L LS CES++R++KGD+Y+D+K A+  S+ A++   +  ++ALKAD
Sbjct: 1   MKKLLSLSLVAISLLNLSACESVDRAIKGDKYIDEKTAKEESEAASKAYEESIQKALKAD 60

Query: 61  KKAFPQLDKAVAKNEAQVLIKTSKGDINIKLFPKYAPLAVENFLTHAKEGYYNGLSFHRV 120
              FPQL K V K EA+V+++TS+GDI +KLFPKYAPLAVENFLTHAK+GYY+ L+FHRV
Sbjct: 61  ASQFPQLTKEVGKEEAKVVMRTSQGDITLKLFPKYAPLAVENFLTHAKKGYYDNLTFHRV 120

Query: 121 IKDFMIQSGDPNGDGTGGKSIWNSKDKKKDSGNGFVNEISPYLYNIRGSLAMANAGADTN 180
           I DFMIQSGDP GDGTGG+SIW  KD KKD+GNGFVNEISP+LY+IRG+LAMANAGA+TN
Sbjct: 121 INDFMIQSGDPKGDGTGGESIWKGKDPKKDAGNGFVNEISPFLYHIRGALAMANAGANTN 180

Query: 181 GSQFFINQSQQDHSKQLSDKKVPKVIIKAYSEGGNPSLDGGYTVFGQVISGMETVDKIAS 240
           GSQF+INQ++++ SK LS    PK II AY  GGNPSLDGGYTVFGQVI GM+ VDKIA+
Sbjct: 181 GSQFYINQNKKNQSKGLSSTNYPKPIISAYEHGGNPSLDGGYTVFGQVIDGMDVVDKIAA 240

Query: 241 VEVTKSDQPKEKITITSIKVIKDYKFK 267
             + ++D+P++ ITITSI ++KDY+FK
Sbjct: 241 TSINQNDKPEQDITITSIDIVKDYRFK 267
```

SEQ ID 18 (GBS205) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 13; MW 31kDa).

GBS205-His was purified as shown in Figure 206, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 11

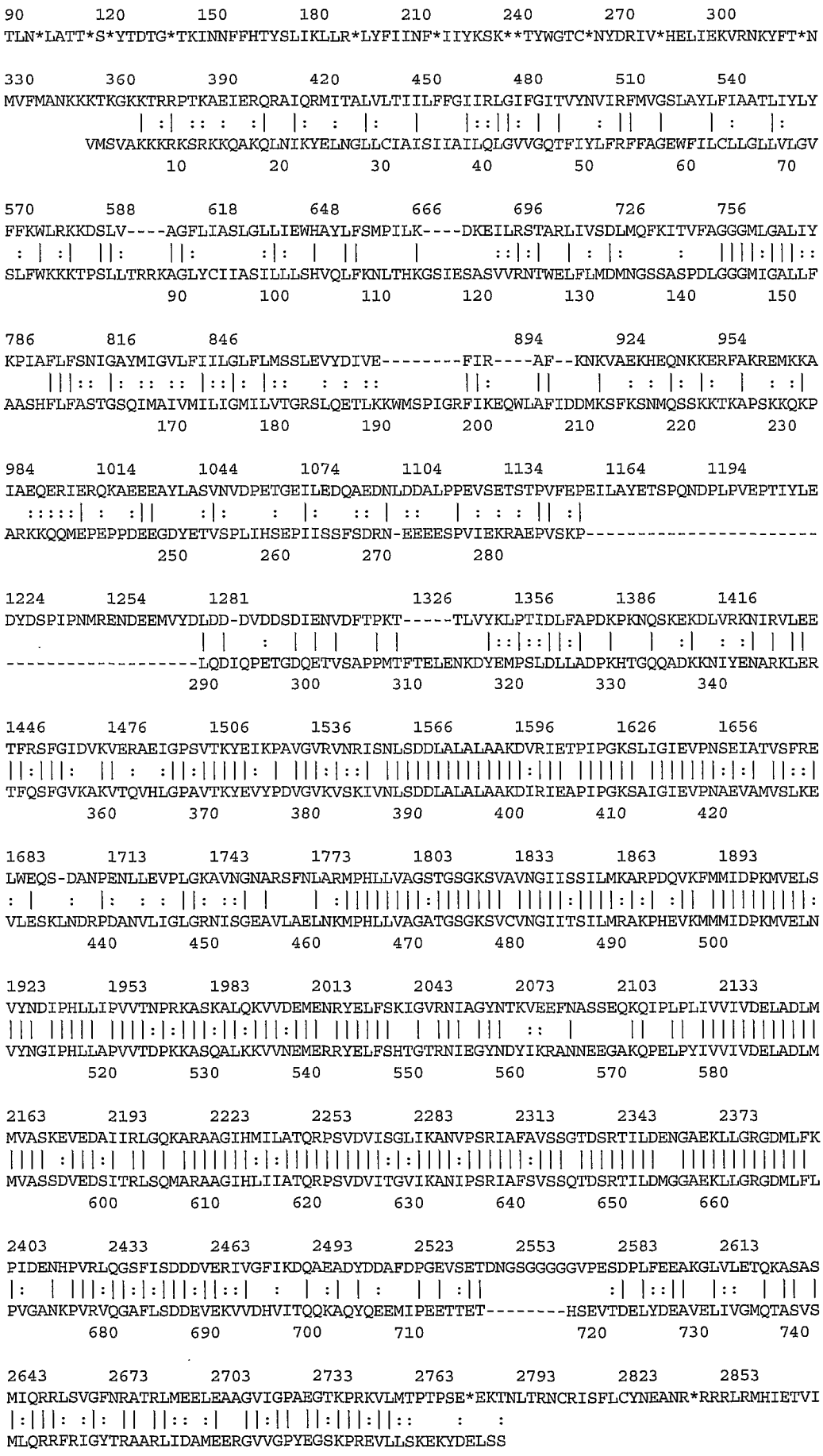
A DNA sequence (GBSx0008) was identified in *S.agalactiae* <SEQ ID 21> which encodes the amino acid sequence <SEQ ID 22>. This protein is predicted to be sporulation protein SpoIIIE (ftsK). Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 10
McG: Discrim Score:    -22.83
GvH: Signal Score (-7.5): -7.13
     Possible site: 39
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 5 value:  -9.24 threshold:  0.0
   INTEGRAL    Likelihood = -9.24   Transmembrane   36 -  52 (  27 -  60)
   INTEGRAL    Likelihood = -9.18   Transmembrane  162 - 178 ( 154 - 188)
   INTEGRAL    Likelihood = -4.04   Transmembrane  597 - 613 ( 595 - 615)
   INTEGRAL    Likelihood = -3.77   Transmembrane   63 -  79 (  58 -  83)
   INTEGRAL    Likelihood = -2.60   Transmembrane   90 - 106 (  88 - 108)
   PERIPHERAL  Likelihood =  1.32     136
 modified ALOM score:   2.35

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4694(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10035> which encodes amino acid sequence <SEQ ID 10036> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB13553 GB:Z99112 DNA translocase [Bacillus subtilis]
 Identities = 352/822 (42%), Positives = 508/822 (60%), Gaps = 70/822 (8%)

Query: 14  KTRRPTKAEIERQRAIQRMITALVLTIILFFGIIRLGIFGITVYNVIRFMVGSLAYLFIA 73
           K +R ++ +  +Q  I+  +  L+   I    I++LG+ G T   + RF  G    L +
Sbjct: 3   KKKRKSRKKQAKQLNIKYELNGLLCIAISIIAILQLGVVGQTFIYLFRFFAGEWFILCLL 62

Query: 74  ATLIYLYFFKWLRKKDSLV----AGFLIASLGLLIEWHAYLFS----MPILKDKEILRST 125
             L+      W +K  SL+    AG       +L+  H  LF        ++   ++R+T
Sbjct: 63  GLLVLGVSLFWKKKTPSLLTRRKAGLYCIIASILLLSHVQLFKNLTHKGSIESASVVRNT 122

Query: 126 ARLIVSDLMQFKITVFAGGGMLGALIYKPIAFLFSNIGAYMIGVLFIILGLFLMSSLEVY 185
             L + D+     +   GGGM+GAL++    FLF++ G+ ++ ++ I++G+ L++   +
Sbjct: 123 WELFLMDMNGSSASPDLGGGMIGALLFAASHFLFASTGSQIMAIVMILIGMILVTGRSLQ 182

Query: 186 DIVE--------FIR----AFKN--KVAEKHEQNKKERFAKREMKKAIAEQERIERQKAE 231
           + ++         FI+    AF +  K  + + Q+ K+  A  + +K   +++++E +  +
Sbjct: 183 ETLKKWMSPIGRFIKEQWLAFIDDMKSFKSNMQSSKKTKAPSKKQKPARKKQQMEPEPPD 242

Query: 232 EEAYLASVNVDPETGEILEDQAEDNLDDALPPEVSETSTPVFEP-EILAYETSPQNDPLP 290
           EE    +V+    +  I+   ++ N ++   P + + + PV +P + +  ET  Q + +
Sbjct: 243 EEGDYETVSPLIHSEPIISSFSDRNEEEE-SPVIEKRAEPVSKPLQDIQPETGDQ-ETVS 300

Query: 291 VEPTIYLEDYDSPIPNMRENDEEMVYDLDDDVDDSDIENVDFTPKTTLVYKLPTIDLFAP 350
             P  + E                                +EN D        Y++P++DL A
Sbjct: 301 APPMTFTE----------------------------LENKD--------YEMPSLDLLAD 324

Query: 351 DKPKNQSKEKDLVRKNIRVLEETFRSFGIDVKVERAEIGPSVTKYEIKPAVGVRVNRISN 410
             K   Q  +K  + +N R LE TF+SFG+  KV +  +GP+VTKYE+ P VGV+V++I N
Sbjct: 325 PKHTGQQADKKNIYENARKLERTFQSFGVKAKVTQVHLGPAVTKYEVYPDVGVKVSKIVN 384

Query: 411 LSDDLALALAAKDVRIETPIPGKSLIGIEVPNSEIATVSFRELWEQS-DANPENLLEVPL 469
           LSDDLALALAAKD+RIE PIPGKS IGIEVPN+E+A VS +E+ E   +  P+  + + L
Sbjct: 385 LSDDLALALAAKDIRIEAPIPGKSAIGIEVPNAEVAMVSLKEVLESKLNDRPDANVLIGL 444

Query: 470 GKAVNGNARSFNLARMPHLLVAGSTGSGKSVAVNGIISSILMKARPDQVKFMMIDPKMVE 529
           G+ ++G A    L +MPHLLVAG+TGSGKSV VNGII+SILM+A+P +VK MMIDPKMVE
Sbjct: 445 GRNISGEAVLAELNKMPHLLVAGATGSGKSVCVNGIITSILMRAKPHEVKMMMIDPKMVE 504

Query: 530 LSVYNDIPHLLIPVVTNPRKASKALQKVVDEMENRYELFSKIGVRNIAGYNTKVEEFNAS 589
           L+VYN IPHLL PVVT+P+KAS+AL+KVV+EME RYELFS  G RNI GYN  ++  N
Sbjct: 505 LNVYNGIPHLLAPVVTDPKKASQALKKVVNEMERRYELFSHTGTRNIEGYNDYIKRANNE 564

Query: 590 SEQKQIPLPLIVVIVDELADLMMVASKEVEDAIIRLGQKARAAGIHMILATQRPSVDVIS 649
              KQ  LP IVVIVDELADLMMVAS +VED+I RL Q ARAAGIH+I+ATQRPSVDVI+
Sbjct: 565 EGAKQPELPYIVVIVDELADLMMVASSDVEDSITRLSQMARAAGIHLIIATQRPSVDVIT 624

Query: 650 GLIKANVPSRIAFAVSSGTDSRTILDENGAEKLLGRGDMLFKPIDENHPVRLQGSFISDD 709
           G+IKAN+PSRIAF+VSS TDSRTILD  GAEKLLGRGDMLF P+  N PVR+QG+F+SDD
Sbjct: 625 GVIKANIPSRIAFSVSSQTDSRTILDMGGAEKLLGRGDMLFLPVGANKPVRVQGAFLSDD 684

Query: 710 DVERIVGFIKDQAEADYDDAFDPGEVSETDNGSGGGGGVPESDPLFEEAKGLVLETQKAS 769
           +VE++V  +  Q +A Y +   P E +ET +           +D L++EA  L++  Q AS
Sbjct: 685 EVEKVVDHVITQQKAQYQEEMIPEETTETHS--------EVTDELYDEAVELIVGMQTAS 736

Query: 770 ASMIQRRLSVGFNRATRLMEELEAAGVIGPAEGTKPRKVLMT 811
            SM+QRR  +G+ RA RL++ +E  GV+GP EG+KPR+VL++
Sbjct: 737 VSMLQRRFRIGYTRAARLIDAMEERGVVGPYEGSKPREVLLS 778

46.5/66.5% over 775aa

  OMNI|NT01BS1964| sporulation protein SpoIIIE Insert characterized

ORF01349(340 - 2733 of 3048)
OMNI|NT01BS1964(6 - 781 of 790) sporulation protein SpoIIIE
%Match = 29.6
%Identity = 46.4  %Similarity = 66.5
Matches = 352  Mismatches = 243  Conservative Sub.s = 152
```



```
                760        770        780        790
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 23> which encodes the amino acid sequence <SEQ ID 24>. Analysis of this protein sequence reveals the following:

```
     Possible site: 51
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.45   Transmembrane   31 -  47 (  25 -  55)
   INTEGRAL    Likelihood = -7.17   Transmembrane  160 - 176 ( 153 - 183)
   INTEGRAL    Likelihood = -4.99   Transmembrane   93 - 109 (  86 - 111)
   INTEGRAL    Likelihood = -4.04   Transmembrane  586 - 602 ( 584 - 604)
   INTEGRAL    Likelihood = -1.22   Transmembrane   64 -  80 (  64 -  80)

----- Final Results -----
               bacterial membrane --- Certainty=0.4779(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
!GB:Z99112 DNA translocase [Bacillus subtilis]     601  e-170
 Identities = 354/816 (43%), Positives = 499/816 (60%), Gaps = 69/816 (8%)

Query: 11  APKKRLTKAEVEKQRAIKRMILSVLMALLLIFAMLRLGVFGVTTYNMIRFLVGSLAYPFM 70
           A KKR ++ +  KQ  IK  +  +L   + I A+L+LGV G T   + RF  G      +
Sbjct: 2   AKKKRKSRKKQAKQLNIKYELNGLLCIAISIIAILQLGVVGQTFIYLFRFFAGEWFILCL 61

Query: 71  FAWLIYLFCFKWLRQKDGMI----AGVVIAFLGLLVEWHAFLFA----MPRMLDQDIFLG 122
              L+      W ++   ++    AG+      +L+  H  LF        +    +
Sbjct: 62  LGLLVLGVSLFWKKKTPSLLTRRKAGLYCIIASILLLSHVQLFKNLTHKGSIESASVVRN 121

Query: 123 TARLITRDLLALRVTEFVGGGMLGALLYKPIAFLFSNIGSYFIGFLFILLGLFLMTPWDI 182
           T  L   D+     +  +GGGM+GALL+    FLF++ GS  +  + IL+G+ L+T   +
Sbjct: 122 TWELFLMDMNGSSASPDLGGGMIGALLFAASHFLFASTGSQIMAIVMILIGMILVTGRSL 181

Query: 183 YD--------VSHFVKEA----VDKLAVAYQENKEKRFIKREEHRLQAEKEALEKQAQEE 230
            +        +  F+KE     +D +  +++ N +    K+ +   + +K A +KQ  E
Sbjct: 182 QETLKKWMSPIGRFIKEQWLAFIDDMK-SFKSNMQSS--KKTKAPSKKQKPARKKQQMEP 238

Query: 231 EKRLAELTVDPETGEIVEDSQSQVSYDLAEDMT-KEPEILAYDSHLKDDETSLFDQ---- 285
           E          E G+          Y+     +  EP I ++     +++E+ + ++
Sbjct: 239 EP-------PDEEGD----------YETVSPLIHSEPIISSFSDRNEEEESPVIEKRAEP 281

Query: 286 --EDLAYAHEEIGAYDSLSALASSEDEMDMDEPVEVDFTPKTHLLYKLPTIDLFAPDKPK 343
             + L     E G  +++SA   +  E++  +              Y++P++DL A  K
Sbjct: 282 VSKPLQDIQPETGDQETVSAPPMTFTELENKD-------------YEMPSLDLLADPKHT 328

Query: 344 NQSKEKNLVRKNIKVLEDTFQSFGIDVKVERAEIGPSVTKYEIKPAVGVRVNRISNLADD 403
            Q  +K  + +N + LE TFQSFG+  KV +  +GP+VTKYE+ P VGV+V++I NL+DD
Sbjct: 329 GQQADKKNIYENARKLERTFQSFGVKAKVTQVHLGPAVTKYEVYPDVGVKVSKIVNLSDD 388

Query: 404 LALALAAKDVRIEAPIPGKSLIGIEVPNSEIATVSFRELWEQS-DANPENLLEVPLGKAV 462
           LALALAAKD+RIEAPIPGKS IGIEVPN+E+A VS +E+ E   +  P+  + + LG+ +
Sbjct: 389 LALALAAKDIRIEAPIPGKSAIGIEVPNAEVAMVSLKEVLESKLNDRPDANVLIGLGRNI 448

Query: 463 NGNARSFNLARMPHLLVAGSTGSGKSVAVNGIISSILMKARPDQVKFMMIDPKMVELSVY 522
           +G A    L +MPHLLVAG+TGSGKSV VNGII+SILM+A+P +VK MMIDPKMVEL+VY
Sbjct: 449 SGEAVLAELNKMPHLLVAGATGSGKSVCVNGIITSILMRAKPHEVKMMMIDPKMVELNVY 508

Query: 523 NDIPHLLIPVVTNPRKASKALQKVVDEMENRYELFSKIGVRNIAGYNTKVEEFNASSEQK 582
           N IPHLL PVVT+P+KAS+AL+KVV+EME RYELFS  G RNI GYN  ++  N     K
Sbjct: 509 NGIPHLLAPVVTDPKKASQALKKVVNEMERRYELFSHTGTRNIEGYNDYIKRANNEEGAK 568

Query: 583 QIPLPLIVVIVDELADLMMVASKEVEDAIIRLGQKARAAGIHMILATQRPSVDVISGLIK 642
           Q  LP IVVIVDELADLMMVAS +VED+I RL Q ARAAGIH+I+ATQRPSVDVI+G+IK
Sbjct: 569 QPELPYIVVIVDELADLMMVASSDVEDSITRLSQMARAAGIHLIIATQRPSVDVITGVIK 628
```

```
Query: 643 ANVPSRMAFAVSSGTDSRTILDENGAEKLLGRGDMLFKPIDENHPVRLQGSFISDDDVER 702
           AN+PSR+AF+VSS TDSRTILD  GAEKLLGRGDMLF P+  N PVR+QG+F+SDD+VE+
Sbjct: 629 ANIPSRIAFSVSSQTDSRTILDMGGAEKLLGRGDMLFLPVGANKPVRVQGAFLSDDEVEK 688

Query: 703 IVNFIKDQTEADYDDAFDPGEVSDNDPGFSGNGGAAEGDPLFEEAKALVLETQKASASMI 762
           +V+ +  Q +A Y +   P E ++     +        D L++EA  L++  Q AS SM+
Sbjct: 689 VVDHVITQQKAQYQEEMIPEETTETHSEVT--------DELYDEAVELIVGMQTASVSML 740

Query: 763 QRRLSVGFNRATRLMDELEEAGVIGPAEGTKPRKVL 798
           QRR  +G+ RA RL+D +EE GV+GP EG+KPR+VL
Sbjct: 741 QRRFRIGYTRAARLIDAMEERGVVGPYEGSKPREVL 776
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 620/818 (75%), Positives = 701/818 (84%), Gaps = 25/818 (3%)

Query: 1   MVFMANKKKTKGKKTRRPTKAEIERQRAIQRMITALVLTIILFFGIIRLGIFGITVYNVI 60
           MV    +KK+  KK  R TKAE+E+QRAI+RMI ++++ ++L F ++RLG+FG+T YN+I
Sbjct: 1   MVKRNQRKKSAPKK--RLTKAEVEKQRAIKRMILSVLMALLLIFAMLRLGVFGVTTYNMI 58

Query: 61  RFMVGSLAYLFIAATLIYLYFFKWLRKKDSLVAGFLIASLGLLIEWHAYLFSMPILKDKE 120
           RF+VGSLAY F+ A LIYL+ FKWLR+KD ++AG +IA LGLL+EWHA+LF+MP + D++
Sbjct: 59  RFLVGSLAYPFMFAWLIYLFCFKWLRQKDGMIAGVVIAFLGLLVEWHAFLFAMPRMLDQD 118

Query: 121 ILRSTARLIVSDLMQFKITVFAGGGMLGALIYKPIAFLFSNIGAYMIGVLFIILGLFLMS 180
           I   TARLI  DL+  ++T F GGGMLGAL+YKPIAFLFSNIG+Y IG LFI+LGLFLM+
Sbjct: 119 IFLGTARLITRDLLALRVTEFVGGGMLGALLYKPIAFLFSNIGSYFIGFLFILLGLFLMT 178

Query: 181 SLEVYDIVEFIRAFKNKVAEKHEQNKKERFAKREMKKAIAEQERIERQKAEEEAYLASVN 240
             ++YD+  F++   +K+A  +++NK++RF KRE  +  AE+E +E+Q  EEE  LA +
Sbjct: 179 PWDIYDVSHFVKEAVDKLAVAYQENKEKRFIKREEHRLQAEKEALEKQAQEEEKRLAELT 238

Query: 241 VDPETGEILEDQAEDNLDDALPPEVSETSTPVFEPEILAYETSPQNDPLPV---EPTIYL 297
           VDPETGEI+ED          +  +++E  T   EPEILAY++  ++D   +   E   Y
Sbjct: 239 VDPETGEIVEDSQSQ-----VSYDLAEDMTK--EPEILAYDSHLKDDETSLFDQEDLAYA 291

Query: 298 ED----YDSPIPNMRENDEEMVYDLDDDVDDSDIENVDFTPKTTLVYKLPTIDLFAPDKP 353
            +    YDS +  +  +++EM  D+D+ V+       VDFTPKT L+YKLPTIDLFAPDKP
Sbjct: 292 HEEIGAYDS-LSALASSEDEM--DMDEPVE------VDFTPKTHLLYKLPTIDLFAPDKP 342

Query: 354 KNQSKEKDLVRKNIRVLEETFRSFGIDVKVERAEIGPSVTKYEIKPAVGVRVNRISNLSD 413
           KNQSKEK+LVRKNI+VLE+TF+SFGIDVKVERAEIGPSVTKYEIKPAVGVRVNRISNL+D
Sbjct: 343 KNQSKEKNLVRKNIKVLEDTFQSFGIDVKVERAEIGPSVTKYEIKPAVGVRVNRISNLAD 402

Query: 414 DLALALAAKDVRIETPIPGKSLIGIEVPNSEIATVSFRELWEQSDANPENLLEVPLGKAV 473
           DLALALAAKDVRIE PIPGKSLIGIEVPNSEIATVSFRELWEQSDANPENLLEVPLGKAV
Sbjct: 403 DLALALAAKDVRIEAPIPGKSLIGIEVPNSEIATVSFRELWEQSDANPENLLEVPLGKAV 462

Query: 474 NGNARSFNLARMPHLLVAGSTGSGKSVAVNGIISSILMKARPDQVKFMMIDPKMVELSVY 533
           NGNARSFNLARMPHLLVAGSTGSGKSVAVNGIISSILMKARPDQVKFMMIDPKMVELSVY
Sbjct: 463 NGNARSFNLARMPHLLVAGSTGSGKSVAVNGIISSILMKARPDQVKFMMIDPKMVELSVY 522

Query: 534 NDIPHLLIPVVTNPRKASKALQKVVDEMENRYELFSKIGVRNIAGYNTKVEEFNASSEQK 593
           NDIPHLLIPVVTNPRKASKALQKVVDEMENRYELFSKIGVRNIAGYNTKVEEFNASSEQK
Sbjct: 523 NDIPHLLIPVVTNPRKASKALQKVVDEMENRYELFSKIGVRNIAGYNTKVEEFNASSEQK 582

Query: 594 QIPLPLIVVIVDELADLMMVASKEVEDAIIRLGQKARAAGIHMILATQRPSVDVISGLIK 653
           QIPLPLIVVIVDELADLMMVASKEVEDAIIRLGQKARAAGIHMILATQRPSVDVISGLIK
Sbjct: 583 QIPLPLIVVIVDELADLMMVASKEVEDAIIRLGQKARAAGIHMILATQRPSVDVISGLIK 642

Query: 654 ANVPSRIAFAVSSGTDSRTILDENGAEKLLGRGDMLFKPIDENHPVRLQGSFISDDDVER 713
           ANVPSR+AFAVSSGTDSRTILDENGAEKLLGRGDMLFKPIDENHPVRLQGSFISDDDVER
Sbjct: 643 ANVPSRMAFAVSSGTDSRTILDENGAEKLLGRGDMLFKPIDENHPVRLQGSFISDDDVER 702

Query: 714 IVGFIKDQAEADYDDAFDPGEVSETDNGSGGGGGVPESDPLFEEAKGLVLETQKASASMI 773
           IV FIKDQ EADYDDAFDPGEVS+ D G  G GG  E DPLFEEAK LVLETQKASASMI
Sbjct: 703 IVNFIKDQTEADYDDAFDPGEVSDNDPGFSGNGGAAEGDPLFEEAKALVLETQKASASMI 762
```

```
Query: 774 QRRLSVGFNRATRLMEELEAAGVIGPAEGTKPRKVLMT 811
           QRRLSVGFNRATRLM+ELE AGVIGPAEGTKPRKVL T
Sbjct: 763 QRRLSVGFNRATRLMDELEEAGVIGPAEGTKPRKVLQT 800
```

SEQ ID 22 (GBS272d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 147 (lane 9; MW 55kDa + lane 10; MW 70kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 147 (lane 11 & 13; MW 85kDa + lane 12; MW 74kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 12

A DNA sequence (GBSx0009) was identified in *S.agalactiae* <SEQ ID 25> which encodes the amino acid sequence <SEQ ID 26>. This protein is predicted to be para-aminobenzoate synthetase (pabB) (pabB). Analysis of this protein sequence reveals the following:

```
Possible site: 61

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4073(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD07357 GB:AE000547 para-aminobenzoate synthetase (pabB)
           [Helicobacter pylori 26695]
 Identities = 204/580 (35%), Positives = 325/580 (55%), Gaps = 50/580 (8%)

Query: 16  YRFKNPTKELIADTLEQVLEVIKEVDYYQSQNYYVVGYLSYEASAAF-DSHFKVSQQKLA 74
           ++++   K+L A  L ++   +  +   +   Y+V GYL YEA  AF D +F+     L
Sbjct: 6   FKYQKSVKKLTATNLNELKNALDFISQNRGNGYFV-GYLLYEARLAFLDENFQSQTPFLY 64

Query: 75  GEHLAY---FTVHKDCENEAFPLSYENVRLADNWTANVSEQEYQEAIANIKGQIRQGNTY 131
            E       +++    E+  +P  +          +++ ++ Y +    +K +++ G+TY
Sbjct: 65  FEQFLERKKYSLEPLKEHAFYPKIH----------SSLDQKTYFKQFKAVKERLKNGDTY 114

Query: 132 QVNYTLELSQQLCSDPFSVYERLMVEQGAGYNAYIAYDDKRILSVSPELFFKKK--DEVL 189
           QVN T++L     + P  V++ ++  Q   + A+I  +   +LS SPELFF+ +  D  +
Sbjct: 115 QVNLTMDLFLDTKAKPKRVFKEVVHNQNTPFKAFIENEFGSVLSFSPELFFELEFLDTAI 174

Query: 190 T--TRPMKGTSARKPTYQEDVAERDWLANDPKNRSENMMIVDLLRNDMGRICDVGTVKVK 247
              T+PMKGT AR      D   R +L ND KNRSEN+MIVDLLRND+ R+    +VKV
Sbjct: 175 KIITKPMKGTIARSKNPLIDEKNRLFLQNDDKNRSENVMIVDLLRNDLSRLALKNSVKVN 234

Query: 248 KLCQVEQYATVWQMTSTIEGVLSPEVTLMSIFQALYPCGSITGAPKISTMAIINELEKRP 307
           +L ++    +V+QM S IE   L  + +L  IF+AL+PCGS+TG PKI TM II  LEKRP
Sbjct: 235 QLFEIISLPSVYQMISEIEAKLPLKTSLFEIFKALFPCGSVTGCPKIKTMQIIESLEKRP 294

Query: 308 RGIYCGTIGLCMPDGQAIFNVPIRTVQMKGQQ--AYYGVGGGITWESQTDSEYEETRQKS 365
           RG+YCG IG+ + + +A+F+VPIRT++ +   + GVG G+T++S+   EYEE+  KS
Sbjct: 295 RGVYCGAIGM-VEEKKALFSVPIRTLEKRVHENFLHLGVGSGVTYKSKAPKEYEESFLKS 353

Query: 366 -AVLTRVNPKFQLITTGRV--TENKLLFSQQ--HVERLVESASYFAYSFDKSKFERELKK 420
             V+ ++  +F++ T ++   + KL  + +  H ERL+ S  YF + +D++  + EL
Sbjct: 354 FFVMPKI--EFEIVETMKIIKKDQKLEINNKNAHKERLMNSTRYFNFKYDENLLDFEL-- 409

Query: 421 YLHQLDEKDYRLKIMLDKTGKVTFEVKQLVNLSKKFLTAEVVVQDYPI-KLSPFTYFKTS 479
                EK+  L+++L+K GK+  E K L  L     + E+ + + PI K + F Y KT+
```

```
Sbjct: 410 ------EKEGVLRVLLNKKGKLIKEYKTLEPLK----SLEIRLSEAPIDKRNDFLYHKTT 459

Query: 480 YRPHIIEGQN--------EKIFVSPEGLLLETSIGNIVLEKNGRFLTPDLSEGGLNGIYR 531
           Y P   + +         ++IF + +  L E +  N+VLE + R LTP  S G LNG
Sbjct: 460 YAPFYQKARALIKKGVMFDEIFYNQDLELTEGARSNLVLEIHNRLLTPYFSAGALNGTGV 519

Query: 532 RHLLKNQKVIEAPLTLKDLESADAIYACNAVRGLYPLNLK 571
             LLK   V  APL L+DL+ A  IY  NA+ GL  + +K
Sbjct: 520 VGLLKKGLVGHAPLKLQDLQKASKIYCINALYGLVEVKIK 559
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 27> which encodes the amino acid sequence <SEQ ID 28>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2669(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 303/572 (52%), Positives = 406/572 (70%), Gaps = 1/572 (0%)

Query: 1   MHIETVIDFKELGKRYRFKNPTKELIADTLEQVLEVIKEVDYYQSQNYYVVGYLSYEASA 60
           MH +T+IDFKELG+RY F  P  EL+A +L+QV  VI++V +YQ   YYVVGYLSYEA+A
Sbjct: 3   MHRKTIIDFKELGQRYLFDEPLVELVAKSLDQVGPVIEKVQHYQQLGYYVVGYLSYEAAA 62

Query: 61  AFDSHFKVSQQKLAGEHLAYFTVHKDCENEAFPLSYENVRLADNWTANVSEQEYQEAIAN 120
            FD+  +    +L  E+LAYFTVHK C+ +  PL Y+++ + + W +   ++ YQ+AI
Sbjct: 63  FFDNALQTHNDRLGNEYLAYFTVHKTCQKKDLPLDYDSITIPNQWVSATQKEAYQKAIET 122

Query: 121 IKGQIRQGNTYQVNYTLELSQQL-CSDPFSVYERLMVEQGAGYNAYIAYDDKRILSVSPE 179
           I  +++QGNTYQVNYTL+L+Q+L  +D  ++Y +L+VEQ AGYNAYIA+D+  ++S SPE
Sbjct: 123 IHREMQQGNTYQVNYTLQLTQELNAADSLAIYNKLVVEQAAGYNAYIAHDEFAVISASPE 182

Query: 180 LFFKKKDEVLTTRPMKGTSARKPTYQEDVAERDWLANDPKNRSENMMIVDLLRNDMGRIC 239
           LFFK++   LTTRPMKGT+ R      D  E DWL  D KNRSENMMIVDLLRNDMG+IC
Sbjct: 183 LFFKQEGNRLTTRPMKGTTKRGVNSWLDQQEHDWLQADGKNRSENMMIVDLLRNDMGKIC 242

Query: 240 DVGTVKVKKLCQVEQYATVWQMTSTIEGVLSPEVTLMSIFQALYPCGSITGAPKISTMAI 299
             G+V+V +LC+VE+Y+TVWQMTSTI G L  +  L+ I +AL+PCGSITGAPK+STMAI
Sbjct: 243 QTGSVRVDRLCEVERYSTVWQMTSTIVGDLKADCDLIDILKALFPCGSITGAPKVSTMAI 302

Query: 300 INELEKRPRGIYCGTIGLCMPDGQAIFNVPIRTVQMKGQQAYYGVGGGITWESQTDSEYE 359
           I  LE +PRGIYCG+IG+C+PDG+  FNVPIRT+Q+   QA YGVGGGITW+S+ + EYE
Sbjct: 303 ITSLEPKPRGIYCGSIGICLPDGRRFFNVPIRTIQLSHNQATYGVGGGITWQSKWEDEYE 362

Query: 360 ETRQKSAVLTRVNPKFQLITTGRVTENKLLFSQQHVERLVESASYFAYSFDKSKFERELK 419
           E  QK+A L R     F L TT +V   K+ F +QH+ RL E+A+YFAY +++   +++L
Sbjct: 363 EVHQKTAFLYRHKQIFDLKTTAKVEHKKIAFLEQHLNRLKEAATYFAYPYNEKALQKQLS 422

Query: 420 KYLHQLDEKDYRLKIMLDKTGKVTFEVKQLVNLSKKFLTAEVVVQDYPIKLSPFTYFKTS 479
            YL   +   YRL I L K GK++   + L  LS  FLTA++ +Q   +  SPFTYFKTS
Sbjct: 423 TYLENKNNAAYRLMIRLSKDGKISLSDQPLEPLSADFLTAQLSLQKKDVTASPFTYFKTS 482

Query: 480 YRPHIIEGQNEKIFVSPEGLLLETSIGNIVLEKNGRFLTPDLSEGGLNGIYRRHLLKNQK 539
           YRPHI +   E++F +  G LLETSIGN+ ++       TP ++ G L G++R+ LL   +
Sbjct: 483 YRPHIEQKSYEQLFYNQAGQLLETSIGNLFVQLGQTLYTPPVAVGILPGLFRQELLATGQ 542

Query: 540 VIEAPLTLKDLESADAIYACNAVRGLYPLNLK 571
             E  +TL DL+ A AI+  NAVRGLYPLNL+
Sbjct: 543 AQEKEVTLADLKEASAIFGGNAVRGLYPLNLE 574
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 13

A DNA sequence (GBSx0010) was identified in *S.agalactiae* <SEQ ID 29> which encodes the amino acid sequence <SEQ ID 30>. Analysis of this protein sequence reveals the following:

```
Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1564(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 31> which encodes the amino acid sequence <SEQ ID 32>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5335(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 220/267 (82%), Positives = 243/267 (90%)

Query: 10  LLLEITKIARATYYYQLKKLNKPNKDKAIKSDIQSIYDEHRGNYGYRRIYLELRNRGFVI 69
           +LLEI  ++R+TYYYQ+K+L + +KD  +K  I+ IYDEH+GNYGYRRI++ELRNRGFV+
Sbjct: 1   MLLEILDLSRSTYYYQVKRLAQGDKDIELKHVIREIYDEHKGNYGYRRIHMELRNRGFVV 60

Query: 70  NHKRVQGLMKSMGLTARIRRKRKYASYKGEVGKKADNLIQRQFEGSKPYEKCYTDVTEFA 129
           NHK+VQ LMK MGL ARIRRKRKY+SYKGEVGKKADNLI+R FEGSKPYEKCYTDVTE A
Sbjct: 61  NHKKVQRLMKVMGLAARIRRKRKYSSYKGEVGKKADNLIKRHFEGSKPYEKCYTDVTELA 120

Query: 130 LPEGKLYLSPVLDGYNSEIIDFTLSRSPDLKQVQTMLERAFPAASYSETILHSDQGWQYQ 189
           LPEGKLYLSPVLDGYNSEIIDFTLSRSP+LKQVQTMLE+ FPA SYS TILHSDQGWQYQ
Sbjct: 121 LPEGKLYLSPVLDGYNSEIIDFTLSRSPNLKQVQTMLEKTFPADSYSGTILHSDQGWQYQ 180

Query: 190 HKSYHQFLEDKGIRPSMSRKGNSPDNGMMESFFGILKSEMFYGLEKSYKSLDDLEQAITD 249
           H+SYH FLE KGI  SMSRKGNSPDNGMMESFFGILKSEMFYGLE +Y+SLD LE+AITD
Sbjct: 181 HQSYHDFLESKGILASMSRKGNSPDNGMMESFFGILKSEMFYGLETTYQSLDKLEEAITD 240

Query: 250 YIFYYNNKRIKAKLKGLSPVQYRTKSF 276
           YIFYYNNKRIKAKLKG SPVQYRTKSF
Sbjct: 241 YIFYYNNKRIKAKLKGFSPVQYRTKSF 267
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 14

A DNA sequence (GBSx0011; GBSx2234) was identified in *S.agalactiae* <SEQ ID 33> which encodes the amino acid sequence <SEQ ID 34>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3578(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 35> which encodes the amino acid sequence <SEQ ID 36>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3869(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 107/170 (62%), Positives = 134/170 (77%)

Query: 1   MKLSYEDKLEIYELRKIGMSWSQISQRYDVRISNLKYMIKLMDRYGVEIVEKGRNEYYPP 60
           MK + E K++IYELR++G S   IS+++D+  S+LKYMI+L+DRYGV IV+K +N YY P
Sbjct: 1   MKFNQETKVKIYELRQMGESIKSISKKFDMAESDLKYMIRLIDRYGVTIVQKCKNHYYSP 60

Query: 61  ELKQEMIDKVLIHGCSQLSVSLDYALSNCSILTNWLSQFKKNGYTIVEKTRGRPSKMGRK 120
           ELKQE+I+KVLI G SQ   SLDYAL   S+L+ W++Q+KKNGYTI+EK RGRPSKMGRK
Sbjct: 61  ELKQEIINKVLIDGQSQKQTSLDYALPTSSMLSRWIAQYKKNGYTILEKPRGRPSKMGRK 120

Query: 121 RKKTWEEMTELERLQEENERLRTENAFLKKLRDLRLRDEALQSERQKQLE 170
           RKK  EEMTE+ERLQ+E E  R ENA LKKLR+ RLRDEA   E+QK  +
Sbjct: 121 RKKNLEEMTEVERLQKELEYPRAENAVLKKLREYRLRDEAKLKEQQKSFK 170
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 15

A DNA sequence (GBSx0012) was identified in *S.agalactiae* <SEQ ID 37> which encodes the amino acid sequence <SEQ ID 38>. This protein is predicted to be oxyR protein. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1323(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10033> which encodes amino acid sequence <SEQ ID 10034> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA91664 GB:Z67753 former trsE (rbcR homolog) [Odontella sinensis]
 Identities = 72/259 (27%), Positives = 127/259 (48%), Gaps = 7/259 (2%)

Query: 5   QKLMYLESIELYSNITKAAAHLFISQPYLSKVIKQLENELEIKLIQSQGHQTFLTYAGQR 64
           Q+L  L++I    + T+AA  LF+SQP LSK IK LE+ L I L+  + +   LT AG+
```

```
Sbjct: 8   QQLRILKAIATEKSFTRAAEVLFVSQPSLSKQIKTLESRLNISLLNRENNIVSLTQAGKL 67

Query: 65  YLFYLKEIDMIERQMAKELYLIRSDKKGEITLGINSGLASSILANVLPKFNLEHPEISVK 124
           +L Y + I  +  +  + L  +++  +G + +G +  + + ++  VL  F   HP+I+++
Sbjct: 68  FLEYSERILALCEESCRVLNDLKTGDRGNLIVGASQTIGTYLMPRVLALFAQNHPQINIE 127

Query: 125 LLENNQNISEQLVASGDIDLAV--GMAPILYKDGIASTTIYRDELFLMIPTTSQLYNAEK 182
           +  ++     + V  GDID+AV  G  P   +  +       DEL L+IP +      +K
Sbjct: 128 VHVDSTRKIAKRVLEGDIDIAVVGGNIPEEIEKNLKVEDFVNDELILIIPKSHPFALKKK 187

Query: 183 RGQIIPFEYPISVLD-NEPLILTPLEYGIGKTIAQFYELHHMSLNQMITTSTVPTAASLS 241
           +       Y ++ +  N    +  L   I   IA F       + Q+ +   + TA SL
Sbjct: 188 KKINKDDLYHLNFITLNSNSTIRKLIDNILIQIA-FEPKQFNIIMQLNSIEAIKTAVSL- 245

Query: 242 LSGMGATFVPQTLIHRYLD 260
             G+GA FV  + I + ++
Sbjct: 246 --GLGAAFVSSSAIEKEIE 262
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 39> which encodes the amino acid sequence <SEQ ID 40>. Analysis of this protein sequence reveals the following:

```
     Possible site: 30
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.28   Transmembrane  109 - 125 ( 109 - 126)
   INTEGRAL    Likelihood = -0.27   Transmembrane  146 - 162 ( 146 - 162)

----- Final Results -----
              bacterial membrane --- Certainty=0.1510(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC22434 GB:U32761 transcriptional regulator [Haemophilus influenzae Rd]
 Identities = 157/303 (51%), Positives = 221/303 (72%)

Query: 2   IRQGESYLDIKQIRYFIAIVENHFNLSQAAELLYVSQPTLSMMINDFEKRENVKLFKRKR 61
           + +G   +DI+ +RYF++IV+N FNLS+A++ LYVSQP LSMMI +FE REN+++FKR
Sbjct: 9   VLRGVKMMDIRHLRYFVSIVDNDFNLSRASQNLYVSQPALSMMITEFENRENIQIFKRAS 68

Query: 62  GRIIGLTYLGDNYYKDAQKVLSLYDDMFLKLHDHSKGLKGSINIGIPPLILSVVFSEVMP 121
           G+IIGLT+ G+NYY+DA++V+  Y+DM   L+      KG+I IGIPPL+LS VFS V+P
Sbjct: 69  GKIIGLTFAGENYYRDAKEVIKRYNDMRTNLYKSKDCKKGTITIGIPPLVLSAVFSSVLP 128

Query: 122 KLILENPGIQFNVKEIGAYQLKNELLVGNVDVAVLLSPTGIADNLVETYEIQRSELSVCL 181
            LIL+NP I F +KEIGAY LK+ELL+  VD+AVLL P  I+ N++++ EI  SEL++ L
Sbjct: 129 HLILKNPDINFIIKEIGAYALKSELLLDKVDLAVLLYPERISKNIIDSIEIHSSELALFL 188

Query: 182 SPRHRLASKKVIQWEDLTDEQLALFDPSFMVHHLVLEACERHQVRPNIILTSSSWDFMLN 241
           SP+H LA K+ I W DL  +++A+FD +FM+HH + EA ER+   P+I+L SS WDF+L+
Sbjct: 189 SPKHVLAKKQQITWADLHQQKMAIFDQTFMIHHHLKEAFERNNCYPDIVLDSSCWDFLLS 248

Query: 242 STKINHNVLTICPKPITELYQLKDIKCIPMERPISWRVVLTRLRKKSYSEIEAYIMDDLL 301
           + K N  +LTI P P+ ELY  K+  C  +E P+ W+V L R RK  Y+ +E YI D LL
Sbjct: 249 AVKTNKELLTILPLPMAELYHSKEFLCRKIESPVPWKVTLCRQRKTVYTHLEEYIFDKLL 308

Query: 302 QSF 304
           ++F
Sbjct: 309 EAF 311
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 61/227 (26%), Positives = 111/227 (48%), Gaps = 10/227 (4%)

Query: 9   YLESIELYSNITKAAAHLFISQPYLSKVIKQLENELEIKLIQ-SQGHQTFLTYAGQRYLF 67
           ++  +E + N+++AA  L++SQP LS +I   E    +KL +  +G    LTY G  Y
Sbjct: 17  FIAIVENHFNLSQAAELLYVSQPTLSMMINDFEKRENVKLFKRKRGRIIGLTYLGDNYYK 76
```

```
Query: 68  YLKEIDMIERQMAKELYLIRSDKKGEITLGINSGLASSILANVLPKFNLEHPEISVKLLE 127
             +++  +   M  +L+      KG I +GI   + S + + V+PK  LE+P I   + E
Sbjct: 77  DAQKVLSLYDDMFLKLHDHSKGLKGSINIGIPPLILSVVFSEVMPKLILENPGIQFNVKE 136

Query: 128 NNQNISEQLVASGDIDLAVGMAPILYKDGIAST-TIYRDELFLMIPTTSQLYNAEKRGQI 186
                + +  G++D+AV ++P    D +  T  I R EL + +    +L  A K+  +
Sbjct: 137 IGAYQLKNELLVGNVDVAVLLSPTGIADNLVETYEIQRSELSVCLSPRHRL--ASKK--V 192

Query: 187 IPFEYPISVLDNEPLILTPLEYGIGKTIAQFYELHHMSLNQMITTST 233
           I +E     L +E L L    + +   + +  E H +  N ++T+S+
Sbjct: 193 IQWE----DLTDEQLALFDPSFMVHHLVLEACERHQVRPNIILTSSS 235
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 16

A DNA sequence (GBSx0013) was identified in *S.agalactiae* <SEQ ID 41> which encodes the amino acid sequence <SEQ ID 42>. This protein is predicted to be aminoacylase (cpsA). Analysis of this protein sequence reveals the following:

```
Possible site: 43

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.75   Transmembrane  385 - 401 ( 385 - 401)

----- Final Results -----
               bacterial membrane --- Certainty=0.1298(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF36227 GB:AF168363 aminoacylase [Lactococcus lactis]
 Identities = 201/395 (50%), Positives = 274/395 (68%), Gaps = 5/395 (1%)

Query: 6   LRHQLFEKLDQKCDQMVAIRRYLHENPELSFKETKTAAYISDFYKGKDCHVQTQFGGMNG 65
           L + L   L Q  ++M+ IRR+LH+ PE+SF+E +T  YI  FYK  DC  +    G  G
Sbjct: 3   LLNNLLTSLTQYENEMIQIRRHLHQYPEISFQEKETFKYIMGFYKELDCEPKLIGKGF-G 61

Query: 66  VVVDIYGDKATDKPIKHIALRADFDALPIQEETGLSFASKTAGVMHACGHDAHTAYLLIL 125
           ++VDI G K+     K +ALRADFDAL I E+  LSF S   GVMHACGHDAHTAYL++L
Sbjct: 62  IIVDIEGGKSG----KTLALRADFDALAIFEDNDLSFKSVNPGVMHACGHDAHTAYLMVL 117

Query: 126 AESLIELKSEFSGHIRILHQPAEEVPPGGAKAMIEAGCLDGIDAVLGIHVMSTMEEGTVQ 185
           A  L+++K E  G +RI+HQPAEEV PGGAK+MI+AG LDG+D ++G+HVM+T++ G +
Sbjct: 118 ARELVKIKQELPGRVRIVHQPAEEVSPGGAKSMIKAGALDGVDNMIGVHVMTTIKTGVIA 177

Query: 186 YHAGPIQTGRATFKVILQGKGGHGSMPHRANDTIVAASSFVMAAQTIVSRRVNPFDTAVV 245
           YH    QTGR+ F + ++G GGH SMP  +ND IVAAS FV   QT++SRR++PFD   V
Sbjct: 178 YHNKETQTGRSNFTITIKGNGGHASMPQLSNDAIVAASYFVTELQTVISRRIDPFDMGTV 237

Query: 246 TIGSFDGKGSANVIKDSVTLEGDVRVMSEETRGVVEEEFKRILDGIAQTYGVSYQLDYQN 305
           TIGSFDG GS N I+D V L+GDVR+M E TR V+ ++ K+I  G+  T+GV   +DY +
Sbjct: 238 TIGSFDGAGSFNAIQDKVLLKGDVRMMKETTRKVIRDQVKQIAKGVGVTFGVEVIVDYDD 297

Query: 306 DYPVLVNNSEVTQKVANSLKSVAIKEILDVIDCDPQTPSEDFAYYAQTIPACFFYVGAHE 365
           +YPVL N+  +T  V +SLK   I E+ +++D  PQ PSEDF+YY Q +P+ FFY+GA
Sbjct: 298 NYPVLFNSENLTHFVVDSLKDQNISEVNNIVDLGPQNPSEDFSYYGQVVPSTFFYIGAQP 357

Query: 366 EGQPYYPHHHPKFQIAESSLMVSAKSMATAALAML 400
           E    YPHH P F++ E S++++AK++AT  +  L
Sbjct: 358 EDGGNYPHHSPLFKMNEKSILIAAKAVATVTINYL 392
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 17

A DNA sequence (GBSx0014) was identified in *S.agalactiae* <SEQ ID 43> which encodes the amino acid sequence <SEQ ID 44>. This protein is predicted to be drug transporter. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:      6.19
GvH: Signal Score (-7.5): -0.899999
     Possible site: 31
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 11 value: -12.15 threshold:  0.0
   INTEGRAL    Likelihood =-12.15   Transmembrane  169 - 185 ( 166 - 190)
   INTEGRAL    Likelihood = -8.86   Transmembrane  229 - 245 ( 224 - 250)
   INTEGRAL    Likelihood = -8.65   Transmembrane   82 -  98 (  78 - 111)
   INTEGRAL    Likelihood = -8.60   Transmembrane  436 - 452 ( 428 - 457)
   INTEGRAL    Likelihood = -7.48   Transmembrane  202 - 218 ( 198 - 222)
   INTEGRAL    Likelihood = -4.99   Transmembrane  334 - 350 ( 332 - 352)
   INTEGRAL    Likelihood = -4.88   Transmembrane  358 - 374 ( 354 - 376)
   INTEGRAL    Likelihood = -4.09   Transmembrane  301 - 317 ( 301 - 317)
   INTEGRAL    Likelihood = -2.81   Transmembrane  102 - 118 ( 101 - 119)
   INTEGRAL    Likelihood = -2.71   Transmembrane   52 -  68 (  50 -  70)
   INTEGRAL    Likelihood = -1.70   Transmembrane  271 - 287 ( 270 - 288)
   PERIPHERAL  Likelihood =  0.32     401
 modified ALOM score:   2.93

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5861(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB02058 GB:Z79702 hypothetical protein Rv2333c [Mycobacterium tuberculosis]
 Identities = 118/405 (29%), Positives = 199/405 (49%), Gaps = 9/405 (2%)

Query: 13  KLLVGIVLAVLSFWLFAQS-ILNMG-PDVQSSLGISSGAMDIGVSSTALFSGLFIVVTGG 70
           +LL  I   +  F +F  + I+N+  PD+Q S  +    +   V+S +L   +FI+
Sbjct: 5   QLLTLIATGLGLFMIFLDALIVNVALPDIQRSFAVGEDGLQWVVASYSLGMAVFIMSAAT 64

Query: 71  LADKLGRVKFTFIGLCLNIIGSLLIVLANGAVLFIMGRIFQGLAAAFIMPSTMALVKTYY 130
           LAD  GR ++  IG+ L  +GS+   LA    +    R  QGL AA +  +++ALV   +
Sbjct: 65  LADLDGRRRWYLIGVSLFTLGSIACGLAPSIAVLTTARGAQGLGAAAVSVTSLALVSAAF 124

Query: 131 -DGKDRQRAVSFWSIGSWGGSGLCSYFGGAVASTLGWRYVFIFSI-IASVVSFLLILGTP 188
            + K++ RA+  W+  +  G+      GG +    GWR +F  ++ + ++V FL +
Sbjct: 125 PEAKEKARAIGIWTAIASIGTTTGPTLGGLLVDQWGWRSIFYVNLPMGALVLFLTLCYVE 184

Query: 189 ESKNVGQKTHFDYLGLIIFIISMLSLNIGISMAQEHGLMNVIPLSLFTVMLIGFVLFYYV 248
           ES N  +   FD  G ++FI+++ +L   +    + G  +V  + +     +G  LF ++
Sbjct: 185 ESCN-ERARRFDLSGQLLFIVAVGALVYAVIEGPQIGWTSVQTIVMLWTAAVGCALFVWL 243

Query: 249 ETRKSNSFIDFHLFENRFY-LGATISNFLLNAVAGTLIVINTYMQQGRQLTPKVAGEMSL 307
           E R SN  +D  LF +  Y L       +  AV G L++    ++Q  R  TP V G M L
Sbjct: 244 ERRSSNPMMDLTLFRDTSYALAIATICTVFFAVYGMLLLTTQFLQNVRGYTPSVTGLMIL 303

Query: 308 GYLVCVLIAIRVGEKILQRFGARKPMLLGAMSTFVGIFLMTLVNIQGPLYLVLVFVGYAL 367
            +   V I   +  ++ R GAR P+L G     +G+ ++     +      LV VG  L
```

```
Sbjct: 304 PFSAAVAIVSPLVGHLVGRIGARVPILAGLCMLMLGLLMLIFSEHRSS---ALVLVGLGL 360

Query: 368 FGTGLGIYATPSTDTAISSIPNEKVGSASGIYKMASSLGGAIGVA 412
           G+G+ +  TP T  A++++P E+ G ASGI     ++G  IG A
Sbjct: 361 CGSGVALCLTPITTVAMTAVPAERAGMASGIMSAQRAIGSTIGFA 405
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 45> which encodes the amino acid sequence <SEQ ID 46>. Analysis of this protein sequence reveals the following:

```
Possible site: 61

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.28   Transmembrane  169 - 185 ( 165 - 189)
   INTEGRAL    Likelihood = -8.23   Transmembrane   12 -  28 (  11 -  32)
   INTEGRAL    Likelihood = -8.17   Transmembrane  429 - 445 ( 423 - 450)
   INTEGRAL    Likelihood = -6.64   Transmembrane  203 - 219 ( 200 - 222)
   INTEGRAL    Likelihood = -5.41   Transmembrane  227 - 243 ( 225 - 245)
   INTEGRAL    Likelihood = -3.72   Transmembrane   82 -  98 (  80 -  99)
   INTEGRAL    Likelihood = -3.72   Transmembrane  136 - 152 ( 135 - 155)
   INTEGRAL    Likelihood = -2.92   Transmembrane  302 - 318 ( 299 - 319)
   INTEGRAL    Likelihood = -2.55   Transmembrane  261 - 277 ( 261 - 277)
   INTEGRAL    Likelihood = -2.07   Transmembrane  331 - 347 ( 331 - 347)
   INTEGRAL    Likelihood = -1.06   Transmembrane   56 -  72 (  56 -  72)
   INTEGRAL    Likelihood = -0.96   Transmembrane  351 - 367 ( 351 - 368)
   INTEGRAL    Likelihood = -0.37   Transmembrane  104 - 120 ( 103 - 120)

----- Final Results -----
              bacterial membrane --- Certainty=0.4312(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
!GB:AJ250422 ORFC [Oenococcus oeni]                271  1e-71
 Identities = 152/445 (34%), Positives = 248/445 (55%), Gaps = 7/445 (1%)

Query: 1   MSHHQQTVSKQTIMAIIAIALIGFSGILSETSMNVTFPTLMSVYQLPLNSLQWMTTIYLL 60
           M    Q VS    +AI+ +A + F G+L ETSMNVTFPTLM  + + LN +QW+TT YLL
Sbjct: 1   MQKDNQPVSLHVKLAILGLAGLAFCGVLIETSMNVTFPTLMQQFSISLNKVQWLTTAYLL 60

Query: 61  AVAIMMTTSATLKKNVRERPLFFMATGLFTFGTILAVLTQSFAIMLLARIFQGIGTGLVM 120
            VA  ++ +A ++K    + +FF A  LF  G I + L  +F I+L+ R+ Q + TGL +
Sbjct: 61  LVAATISIAAFIEKRFIFKKIFFWAGLLFIIGVICSALAPNFLILLIGRLIQALSTGLAI 120

Query: 121 PQMFNIILERVPMHKVGLFMGFAGLIISLAPAFGPTYGGFMISHFSWQWIFICILPVPLI 180
           P +   I++++P  K G +M     ++   P+ GPTYGG +    SW+ IF  +LP+ LI
Sbjct: 121 PLLITEIMQQIPQKKQGSYMELVEWLLLWQPSLGPTYGGVITQDLSWRLIFWFVLPIGLI 180

Query: 181 AGILAYYYLEDSPVSEKVPFDWLAFIALSISLTSALLAITSLE-NGSVNLYYLGLFILSF 239
           A ++   ++E      K+PF W  FI+L ++L S  +A+ +    G  ++ + G  +++
Sbjct: 181 AWLIGLSFIEQKSSPSKIPFAWKQFISLILALLSITVAVNNAGIYGWTSIKFYGFLLIAV 240

Query: 240 IL---FLYKNLTAKQPFLDIRILKIPSLTFGLIPFFVFQLINLGINFLTPNFIVMEKIAN 296
           IL   F+  +  ++Q  + I I K       L+ +F+ Q I L + FL PN+  +
Sbjct: 241 ILLIVFIKLSTNSRQALISISIFKKWEFVCPLLIYFLIQFIQLSLTFLLPNYAQLILKKG 300

Query: 297 SSQAGMVLLPGTLLGALLAPAFGKLYDQKGARLSLYLGNALFSLSLIIMTLQTRHFMLLP 356
              +G++LL G+L+ A+L P  G++ D    ++ L +G      S I  T+  R+  +
Sbjct: 301 VMISGIMLLCGSLISAILQPLTGRMLDSFSVKIPLVIGAFFLITSTISFTIFQRYLSVFL 360

Query: 357 FTLLYILFTFGRNMGFNNSLATAIRELPAEKNADATAIFQMMQQFAGALGTAMAS-LIAN 415
              LY+++  G +  FNNSL  A+++LP +  +D  A+F  +QQ+AG+LGT++AS L+AN
Sbjct: 361 IAALYVIYMIGFSFVFNNSLTYALQKLPLKLISDGNAVFNTLQQYAGSLGTSVASALLAN 420

Query: 416 SQAEFTSGVQSVYLLFTIFALLDFI 440
              T G QS Y       +L+FI
Sbjct: 421 GIG--TDGKQSNYTGSRHIFILNFI 443
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 91/369 (24%), Positives = 160/369 (42%), Gaps = 14/369 (3%)

Query: 82  FIGLCLNIIGSLLIVLANGAVLFIMGRIFQGLAAAFIMPSTMALVKTYYDGKDRQRAVSF 141
           F+   L   G++L VL     + ++ RIFQG+    +MP    ++            + F
Sbjct: 83  FMATGLFTFGTILAVLTQSFAIMLLARIFQGIGTGLVMPQMFNIILERVPMHKVGLFMGF 142

Query: 142 WSIGSWGGSGLCSYFGGAVASTLGWRYVFIFSIIASVVSFLLILGTPESKNVGQKTHFDY 201
             +           +GG + S   W+++FI  +   +++ +L     E   V +K  FD+
Sbjct: 143 AGLIISLAPAFGPTYGGFMISHFSWQWIFICILPVPLIAGILAYYYLEDSPVSEKVPFDW 202

Query: 202 LGLIIFIISMLSLNIGISMAQEHGLMNVIPLSLFTVMLIGFVLFYYVETRKSNSFIDFHL 261
           L  I   IS+ S  + I+ + E+G +N+  L LF   ++ F+LF Y        F+D  +
Sbjct: 203 LAFIALSISLTSALLAIT-SLENGSVNLYYLGLF---ILSFILFLYKNLTAKQPFLDIRI 258

Query: 262 FENRFYLGATISNFLLNAV-AGTLIVINTYMQQGRQLTPKVAGEMSL-GYLVCVLIAIRV 319
            +        I  F+   + G   +   ++   +      AG + L G L+  L+A
Sbjct: 259 LKIPSLTFGLIPFFVFQLINLGINFLTPNFIVMEKIANSSQAGMVLLPGTLLGALLAPAF 318

Query: 320 GEKILQRFGARKPMLLGAMSTFVGIFLMTLVNIQGPLYLVLVF-VGYALFGTGLGIYATP 378
           G K+  + GAR  + LG     + + +MTL   Q   +++L F + Y LF  G   +
Sbjct: 319 G-KLYDQKGARLSLYLGNALFSLSLIIMTL---QTRHFMLLPFTLLYILFTFGRNMGFNN 374

Query: 379 STDTAISSIPNEKVGSASGIYKMASSLGGAIGVATSIAIYHAFSGNADFHKAALCGLILN 438
           S  TAI  +P EK   A+ I++M     GA+G A +  I ++    A+F        +L
Sbjct: 375 SLATAIRELPAEKNADATAIFQMMQQFAGALGTAMASLIANS---QAEFTSGVQSVYLLF 431

Query: 439 LVFCSLSIL 447
            +F  L  +
Sbjct: 432 TIFALLDFI 440
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 18

A DNA sequence (GBSx0015) was identified in *S.agalactiae* <SEQ ID 47> which encodes the amino acid sequence <SEQ ID 48>. This protein is predicted to be transposase. Analysis of this protein sequence reveals the following:

```
Possible site: 45

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3116(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 19

A DNA sequence (GBSx0016) was identified in *S.agalactiae* <SEQ ID 49> which encodes the amino acid sequence <SEQ ID 50>. This protein is predicted to be L11 protein (rplK). Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1859(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA53739 GB:X76134 L11 protein [Staphylococcus carnosus]
 Identities = 117/139 (84%), Positives = 129/139 (92%)

Query: 1   MAKKVEKLVKLQIPAGKATPAPPVGPALGQAGINIMGFTKEFNARTADQAGMIIPVVISV 60
           MAKKVEK+VKLQIPAGKA PAPPVGPALGQAG+NIMGF KEFNART +QAG+IIPV ISV
Sbjct: 1   MAKKVEKVVKLQIPAGKANPAPPVGPALGQAGVNIMGFCKEFNARTQEQAGLIIPVEISV 60

Query: 61  YEDKSFDFITKTPPAAVLLKKAAGVEKGSGEPNKTKVATITRAQVQEIAETKMPDLNAAN 120
           YED+SF FITKTPPA VLLKKAAGVEKGSGEPNK KVAT+T+ QV+EIA+TKMPDLNAA+
Sbjct: 61  YEDRSFTFITKTPPAPVLLKKAAGVEKGSGEPNKNKVATVTKDQVREIAQTKMPDLNAAD 120

Query: 121 LESAMRMIEGTARSMGFTV 139
            E+AMR+IEGTARSMG TV
Sbjct: 121 EEAAMRIIEGTARSMGITV 139
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 51> which encodes the amino acid sequence <SEQ ID 52>. Analysis of this protein sequence reveals the following:

```
Possible site: 45

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4276(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 136/141 (96%), Positives = 139/141 (98%)

Query: 1   MAKKVEKLVKLQIPAGKATPAPPVGPALGQAGINIMGFTKEFNARTADQAGMIIPVVISV 60
           MAKKVEKLVKLQIPAGKATPAPPVGPALGQAGINIMGFTKEFNARTADQAGMIIPVVISV
Sbjct: 25  MAKKVEKLVKLQIPAGKATPAPPVGPALGQAGINIMGFTKEFNARTADQAGMIIPVVISV 84

Query: 61  YEDKSFDFITKTPPAAVLLKKAAGVEKGSGEPNKTKVATITRAQVQEIAETKMPDLNAAN 120
           YEDKSFDFITKTPPAAVLLKKAAGVEKGSG PN TKVAT+TRAQVQEIAETKMPDLNAAN
Sbjct: 85  YEDKSFDFITKTPPAAVLLKKAAGVEKGSGTPNTTKVATVTRAQVQEIAETKMPDLNAAN 144

Query: 121 LESAMRMIEGTARSMGFTVTD 141
           +E+AMRMIEGTARSMGFTVTD
Sbjct: 145 IEAAMRMIEGTARSMGFTVTD 165
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 20

A DNA sequence (GBSx0017) was identified in *S.agalactiae* <SEQ ID 53> which encodes the amino acid sequence <SEQ ID 54>. This protein is predicted to be ribosomal protein L1 (rplA). Analysis of this protein sequence reveals the following:

```
Possible site: 30

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2285(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB11879 GB:Z99104 ribosomal protein L1 (BL1) [Bacillus subtilis]
 Identities = 144/228 (63%), Positives = 177/228 (77%)

Query: 1   MAKKSKNLRAALEKIDSTKAYSVEEAVALAKETNFAKFDATVEVSYNLNIDVKKADQQIR 60
           MAKK K    A + +D +KAY V EAVAL K+TN AKFDATVEV++ L +D  K  QQIR
Sbjct: 1   MAKKGKKYVEAAKLVDHSKAYDVSEAVALVKKTNTAKFDATVEVAFRLGVDPSKNHQQIR 60

Query: 61  GAMVLPAGTGKTSRVLVFARGAKAEEAKAAGADFVGEDDLVAKIQGGWLDFDVVIATPDM 120
           GA+VLP GTGKT RVLVFA+G KA+EA+AAGADFVG+ D + KIQ GW DFDV++ATPDM
Sbjct: 61  GAVVLPNGTGKTQRVLVFAKGEKAKEAEAAGADFVGDTDYINKIQQGWFDFDVIVATPDM 120

Query: 121 MALVGRLGRVLGPRNLMPNPKTGTVTMDVAKAVEESKGGKITYRADKAGNVQALIGKVSF 180
           M  VG++GRVLGP+ LMPNPKTGTVT +V KA+ E K GK+ YR DKAGN+   IGKVSF
Sbjct: 121 MGEVGKIGRVLGPKGLMPNPKTGTVTFEVEKAIGEIKAGKVEYRVDKAGNIHVPIGKVSF 180

Query: 181 DDAKLVDNFKAFNDVIVKAKPATAKGTYITNLSITTTQGVGIKVDPNS 228
           +D KLV+NF    D I+KAKPA AKG Y+ N+++T+T G G+KVD ++
Sbjct: 181 EDEKLVENFTTMYDTILKAKPAAAKGVYVKNVAVTSTMGPGVKVDSST 228
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 55> which encodes the amino acid sequence <SEQ ID 56>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2309(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 208/229 (90%), Positives = 220/229 (95%)

Query: 1   MAKKSKNLRAALEKIDSTKAYSVEEAVALAKETNFAKFDATVEVSYNLNIDVKKADQQIR 60
           MAKKSK +RAALEK+DSTKAYSVEEAVAL KETNFAKFDA+VEV+YNLNIDV+KADQQIR
Sbjct: 1   MAKKSKQMRAALEKVDSTKAYSVEEAVALVKETNFAKFDASVEVAYNLNIDVRKADQQIR 60

Query: 61  GAMVLPAGTGKTSRVLVFARGAKAEEAKAAGADFVGEDDLVAKIQGGWLDFDVVIATPDM 120
           GAMVLP GTGKT RVLVFARGAKAEEAKAAGADFVGEDDLVAKI GGWLDFDVVIATPDM
Sbjct: 61  GAMVLPNGTGKTQRVLVFARGAKAEEAKAAGADFVGEDDLVAKINGGWLDFDVVIATPDM 120

Query: 121 MALVGRLGRVLGPRNLMPNPKTGTVTMDVAKAVEESKGGKITYRADKAGNVQALIGKVSF 180
           MA+VGRLGRVLGPRNLMPNPKTGTVTMDVAKAVEESKGGKITYRADKAGNVQALIGKVSF
Sbjct: 121 MAIVGRLGRVLGPRNLMPNPKTGTVTMDVAKAVEESKGGKITYRADKAGNVQALIGKVSF 180

Query: 181 DDAKLVDNFKAFNDVIVKAKPATAKGTYITNLSITTTQGVGIKVDPNSL 229
           D  KLV+NFKAF+DV+ KAKPATAKGTY+ N+SIT+TQGVGIKVDPNSL
```

```
Sbjct: 181 DADKLVENFKAFHDVMAKAKPATAKGTYMANVSITSTQGVGIKVDPNSL 229
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 21

A DNA sequence (GBSx0018) was identified in *S.agalactiae* <SEQ ID 57> which encodes the amino acid sequence <SEQ ID 58>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> May be a lipoprotein

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10029> which encodes amino acid sequence <SEQ ID 10030> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB04286 GB:AP001509 nickel transport system (nickel-binding
          protein) [Bacillus halodurans]
 Identities = 209/541 (38%), Positives = 324/541 (59%), Gaps = 14/541 (2%)

Query: 5   RRNILLSITCLLMVTLTACHSQDS----KSHKLNSDK-LTLAWGEDFGDVNPHRYNPDQF 59
           R+ ILL +  L+   L  C   +S       + N++K +T +W  D G +NPH YNP Q
Sbjct: 6   RKLILLFVISLISSILVGCAESESGTVSNEGEENTEKSITFSWPRDIGPMNPHVYNPSQL 65

Query: 60  VIQDMVYEGLVRYGDNGKIEPALAKSWSISQDGKTYTFKLRNA-KYSDGSNFNAANVKRN 118
             Q M+YE LV Y + G+++P LA SW+IS+DGK YTFKLR   ++SDG+ FNA  VK+N
Sbjct: 66  FAQSMIYEPLVSYTEGGELQPHLADSWTISEDGKEYTFKLREGVQFSDGTPFNAEIVKKN 125

Query: 119 FDSIFSKSNRGNHNWFNLTNQLENYRALNQSTFEIKLKQAYSATLYDLSMIRPIRFLSDS 178
           FD+    S+   H+W  + N LE    +++ TF++ LK+ Y   L DL+++RP+RFL ++
Sbjct: 126 FDTWIEHSSL--HSWLGVMNVLEKTEVVDEFTFKMVLKEPYYPALQDLAVVRPVRFLGEA 183

Query: 179 AFPKGDDTTKKNVKKPIGTGQWVVKSKKQNEYITFKRNENYWGKKPKLKEVTVKVIPDAQ 238
            FP   DT++  +K+PIGTG W++   KQ+EY  F RN NYWG+ PK+ +VTVK+IPDA+
Sbjct: 184 GFPDDGDTSQ-GIKEPIGTGPWMLSDYKQDEYAVFTRNPNYWGESPKIDKVTVKIIPDAE 242

Query: 239 TRALAFESGDVDLIYGNGIIGLDTFAQYTKDKKYVTAISQPMSTRLLLLNAKESIFQDKK 298
           TR LAFESG++DLI+G G+I +D F Q  +  +Y T +S+P+ TR LLLN       D +
Sbjct: 243 TRVLAFESGELDLIFGEGVISMDAFNQLKESGQYGTDLSEPVGTRSLLLNTSNEKLADLR 302

Query: 299 VRQAMNHAIDKVSIAKNTFRGTEKPADTIFSKSTSHSDAKLNPYSYNVDKANQLLDQAGW 358
           VR A++H  +K ++ +    G E+ AD I S +  ++D  + P  Y+V++AN  LD+AGW
Sbjct: 303 VRLALHHGFNKQAMVEGVTLGLEEKADNILSTNFPYTDIDVEPIEYDVEQANAYLDEAGW 362

Query: 359 KMGKDK-VREKDGKTLTLRLPYIATKATDKDLVTYFQGEWRKIGINVSLIAMEEDDYWAN 417
           ++   K VREK+G+ L L L Y  T    K +    Q EW  IG+ + +  +E
Sbjct: 363 ELPAGKTVREKNGEQLELELIYDKTDPLQKAMAETMQAEWAAIGVKLDITGLELTTQIQR 422

Query: 418 AKKGNFDMMLTYSWGAPWDPHAWMSALTAKADHGHPENIALENLATKTEMDRLIKSALVD 477
            + G+FD+   Y++GAP+DPH++++ + A+A  G  E  A  NL+ K E+D  ++ L
Sbjct: 423 RRAGDFDVDFWYNYGAPYDPHSFIN-VVAEAGWGVAE--AHSNLSMKEELDEQVRATLAS 479

Query: 478 PKEENVDRDYKKVLELLHDEAVYIPLTYQSVISVYRKGDFKTMRFAPEENSFPLRYIEKNN 538
             E      Y  +L  L +++V++P++Y     VY++ +     F    +  P   I+ +N
Sbjct: 480 TDETERQELYGSILNTLQEQSVFVPISYIKKTVVYQE-NVNEFIFPANRDEHPFNGIDVSN 539
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 59> which encodes the amino acid sequence <SEQ ID 60>. Analysis of this protein sequence reveals the following:

```
     Possible site: 24

>>> May be a lipoprotein

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 131/497 (26%), Positives = 220/497 (43%), Gaps = 55/497 (11%)

Query: 8   ILLSITCLLMVTLTACHSQDSKSHKLN-----SDKLTLAWGEDFGDVNPHRYNP-DQFVI 61
           I L +T L++V   AC  Q  ++ +        D+L ++ G       PH ++P D++ +
Sbjct: 13  ITLFLTGLILV---ACQQQKPQTKERQRKQRPKDELVVSMGAKL----PHEFDPKDRYGV 65

Query: 62  QD---MVYEGLVRYGDNGKIEPALAKSWSISQDGKTYTFKLRNA-KYSDGSNFNAANVKR 117
            +   + +  L++       I+  LAK++ +S+DG T++F L +  K+S+G    A +VK
Sbjct: 66  HNEGNITHSTLLKRSPELDIKGELAKTYHLSEDGLTWSFDLHDDFKFSNGEPVTADDVKF 125

Query: 118 NFDSIFSKSNRGNHNWFNLTNQLENYRALNQSTFEIKLKQAYSATLYDLSMIRPIRFLSD 177
            +D +     + +   ++LT  ++N   + ++   I L +A+S     L+ I PI
Sbjct: 126 TYDML-----KADGKAWDLTF-IKNVEVVGKNQVNIHLTEAHSTFTAQLTEI-PI----- 173

Query: 178 SAFPKG--DDTTKKNVKKPIGTGQWVVKSKKQNEYITFKRNENYWGKKPKLKEVTVKVIP 235
               PK   +D  K N   PIG+G ++VK  K  E   F RN  + GKKP  K+ T  V+
Sbjct: 174 --VPKKHYNDKYKSN---PIGSGPYMVKEYKAGEQAIFVRNPYWHGKKPYFKKWT-WVLL 227

Query: 236 DAQTRALAFESGDVDLIYGNGIIGLDTFAQYTK----DKKYVTAISQPMSTRLLLLNAKE 291
           D  T   A ESGDVD+IY    +  D   + T+        V  +S P   + ++ ++ +
Sbjct: 228 DENTALAALESGDVDMIYATPELA-DKKVKGTRLLDIPSNDVRGLSLPYVKKGVITDSPD 286

Query: 292 ------SIFQDKKVRQAMNHAIDKVSIAKNTFRGTEKPADTIFSKSTSHSDAKLNPYSYN 345
                 +  D  +R+A+   +++  +      G  KPA +I  K T   + K
Sbjct: 287 GYPVGNDVTSDPAIRKALTIGLNRQKVLDTVLNGYGKPAYSIIDK-TPFWNPKTAIKDNK 345

Query: 346 VDKANQLLDQAGWKMGKDKVREKDGKTLTLRLPYIATKATDKDLVTYFQGEWRKIGINVS 405
           V KA QLL +AGWK   D  R+K        L Y        +L      + + +GI +
Sbjct: 346 VAKAKQLLTKAGWKEQADGSRKKGDLDAAFDLYYPTNDQLRANLAVEVAEQAKALGITIK 405

Query: 406 LIAMEEDDYWANAKKGNFDMMLTYSWGAPWDPHAWMSALTAKADHGHPENIALENLATKT 465
           L A      W      + D  L Y+ G       + S   + A  G   NI   N  T T
Sbjct: 406 LKASN----WDEMATKSHDSALLYAGGRHHAQQFYESHHPSLAGKGW-TNITFYNNPTVT 460

Query: 466 E-MDRLIKSALVDPKEE 481
           + +D+ + S+ +D   E
Sbjct: 461 KYLDKAMTSSDLDKANE 477
```

A related GBS gene <SEQ ID 8469> and protein <SEQ ID 8470> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: 22   Crend: 5
McG: Discrim Score:      7.69
GvH: Signal Score (-7.5): -3.34
     Possible site: 25
>>> May be a lipoprotein
ALOM program   count: 0 value:   7.21 threshold:  0.0
   PERIPHERAL  Likelihood =  7.21     273
 modified ALOM score:  -1.94

*** Reasoning Step: 3
```

```
----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
                                                                                                    Escherichia coli
  EGAD|8250|   nickel-binding  periplasmic  protein  precursor  Insert  characterized
OMNI|NT01EC4139 oligopeptide transporter putative substrate binding
 domain, putative Insert characterized
  SP|P33590|NIKA_ECOLI NICKEL-BINDING PERIPLASMIC PROTEIN PRECURSOR. Edit characterized
  GP|404845|emb|CAA51659.1||X73143 NikA Insert characterized
  GP|466612|gb|AAB18451.1||U00039 nikA Insert characterized
  GP|1789887|gb|AAC76501.1||AE000423  periplasmic  binding  protein  for  nickel  Insert
characterized
  PIR|S39594|S39594  nickel-binding  periplasmic  protein  precursor  -  Escheri  Insert
characterized

ORF02080(391 - 1905 of 2223)
EGAD|8250|EC3476(21 - 520 of 524) nickel-binding periplasmic protein precursor {Escherichia
coli}OMNI|NT01EC4139  oligopeptide  transporter  putative  substrate  binding  domain,
putativeSP|P33590|NIKA_ECOLI  NICKEL-BINDING  PERIPLASMIC  PROTEIN
PRECURSOR.GP|404845|emb|CAA51659.1||X73143  NikA  {Escherichia
coli}GP|466612|gb|AAB18451.1||U00039  nikA  {Escherichia
coli}GP|1789887|gb|AAC76501.1||AE000423 periplasmic binding protein for nickel {Escherichia
coli}PIR|S39594|S39594 nickel-binding periplasmic protein precursor - Escheri
%Match = 26.9
%Identity = 41.3  %Similarity = 63.7
Matches = 208  Mismatches = 175  Conservative Sub.s = 113

147       177       207       237       267       297       327       357
SP*IIDTYTLSQSVYSHNFLLRRMQNQYNVGNTSSVDYHKLXX*LIXXXCLKK*LTKLKRKLVKMRRNILLSITCLLMVT

                                                                    MLSTLRRTL

387       417       447       477       507       537       567       597
LTACHSQDSKSHKLNSDKLTLAWGEDFGDVNPHRYNPDQFVIQDMVYEGLVRYGDNGKIEPALAKSWSISQDGKTYTFKL
:          |    |::| ||  : | :||| | |:|   | |||| ||:|  :| : | |||||: |:||||:|| |
FALLACASFIVHAAAPDEITTAWPVNVGPLNPHLYTPNQMFAQSMVYEPLVKYQADGSVIPWLAKSWTHSEDGKTWTFTL
         20        30        40        50        60        70        80

624       654       684       714       744       774       804       834
RN-AKYSDGSNFNAANVKRNFDSIFSKSNRGNHNWFNLTNQLENYRALNQSTFEIKLKQAYSATLYDLSMIRPIRFLSDS
|:  |:|:|  |:|     || :::   ||  | |: | ||: : :||::: ::| || ||   | :|:: || ||:: |
RDDVKFSNGEPFDAEAAAENFRAVL--DNRQRHAWLELANQIVDVKALSKTELQITLKSAYYPFLQELALPRPFRFIAPS
        100       110       120       130       140       150       160

864       894       924       954       984      1014      1044      1071
AFPKGDDTTKKNVKKPIGTGQWVVKSKKQNEYITFKRNENYWGKKPKLKEVTVKVIPDAQTRALAFESGDVDLIYGN-GI
 |   :  |   :| |||||  |::::  | |:|  | |||||||:|| :|::|  ||||  |||:|||:||:||:||| |:
QF--KNHETMNGIKAPIGTGPWILQESKLNQYDVFVRNENYWGEKPAIKKITFNVIPDPTTRAVAFETGDIDLLYGNEGL
            180       190       200       210       220       230       240

1101      1131      1161      1191      1221      1251      1281      1311
IGLDTFAQYTKDKKYVTAISQPMSTRLLLLNAKESIFQDKKVRQAMNHAIDKVSIAKNTFRGTEKPADTIFSKSTSHSDA
: |||||:::::  | | :|||: | :| ||  ::   :  ||:|:|:|::| |:  | : ||:: |||:|: |  :::
LPLDTFARFSQNPAYHTQLSQPIETVMLALNTAKAPTNELAVREALNYAVNKKSLIDNALYGTQQVADTLFAPSVPYANL
            260       270       280       290       300       310       320

1341      1371      1395      1425      1455      1485      1515      1545
KLNPYSYNVDKANQLLDQAGWKM--GKDKVREKDGKTLTLRLPYIATKATDKDLVTYFQGEWRKIGINVSLIAMEEDDYW
 | |  |:  ||  ||::||| :  ||| :|||:|: | : | :| | |  | :    | : |:|| :|||  ||   :
GLKPSQYDPQKAKALLEKAGWTLPAGKD-IREKNGQPLRIELSFIGTDALSKSMAEIIQADMRQIGADVSLIGEEESSIY
            340       350       360       370       380       390       400

1575      1605      1635      1665      1695      1725      1755      1785
ANAKKGNFDMMLTYSWGAPWDPHAWMSALTAKADHGHPENIALENLATKTEMDRLIKSALVDPKEENVDRDYKKVLELLH
```

```
|  :  |  |  |::  :||||:||||::|::       |  :  |  :  ||  |  :|:  |    |     |       |: :|  ||
ARQRDGRFGMIFHRTWGAPYDPHAFLSSM---RVPSHADFQAQQGLADKPLIDKEIGEVLATHDETQRQALYRDILTRLH
             420        430          440        450        460        470        480

1815      1845      1875      1905      1935      1965      1995      2025
DEAVYIPLTYQSVISVYRKGDFKTMRFAPEENSFPLRYIEKNNVSK*FDHQKNIVSFFGIVFHITSNIYSYQTINS*FSR
|||||:|::|  |::  |   |  ::   : :||      |:   |:
DEAVYLPISYISMMVV-SKPELGNIPYAPIATEIPFEQIKPVKP
                  500        510        520
```

There is also homology to SEQ ID 318. An alignment of the GAS and GBS sequences follows:

```
Identities = 44/186 (23%), Positives = 78/186 (41%), Gaps = 27/186 (14%)

Query: 65  VITQMV-DGLLENDEYGNLVPSLAKDWKVSKDGLTYTYTLRDGVSWYTADGEEYAPVTAE 123
           VI  MV +GL+   + G + P+LAK W +S+DG TYT+ LR+      +DG  +     +
Sbjct: 57  VIQDMVYEGLVRYGDNGKIEPALAKSWSISQDGKTYTFKLRNA---KYSDGSNFNAANVK 113

Query: 124 DFVTGLKHAVDDKSDALYVVEDSIKNLKAYQNGEVDFKEVGVKALDDKTVQYTLNKPESY 183
                +   +  +   + + + ++N                +AL+  T +  L   ++Y
Sbjct: 114 RNFDSIFSKSNRGNHNWFNLTNQLEN---------------YRALNQSTFEIKLK--QAY 156

Query: 184 WNSKTTYSVLFPVNAKFLKS----KGKDFGTTDPSSILVNGAYFLSAFTSKSSMEFHKNE 239
             S T Y +      +FL      KG D    +    +  G + + +      + F +NE
Sbjct: 157 --SATLYDLSMIRPIRFLSDSAFPKGDDTTKKNVKKPIGTGQWVVKSKKQNEYITFKRNE 214

Query: 240 NYWDAK 245
           NYW  K
Sbjct: 215 NYWGKK 220
```

SEQ ID 8470 (GBS186) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 35 (lane 7; MW 60kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 41 (lane 6; MW 85.7kDa).

GBS186-GST was purified as shown in Figure 202, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 22

A DNA sequence (GBSx0019) was identified in *S.agalactiae* <SEQ ID 61> which encodes the amino acid sequence <SEQ ID 62>. Analysis of this protein sequence reveals the following:

```
Possible site: 37

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -5.95   Transmembrane  101 - 117 (  99 - 123)
   INTEGRAL    Likelihood = -4.73   Transmembrane  276 - 292 ( 275 - 293)
   INTEGRAL    Likelihood = -1.12   Transmembrane  232 - 248 ( 232 - 248)
   INTEGRAL    Likelihood = -0.96   Transmembrane  151 - 167 ( 150 - 169)

----- Final Results -----
              bacterial membrane --- Certainty=0.3378(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB04287 GB:AP001509 nickel transport system (permease)
          [Bacillus halodurans]
 Identities = 119/304 (39%), Positives = 174/304 (57%)

Query: 5   SSIIKKILSAFLALFFISLLTFILIKLSTVNSAENYLRLSKISVSPEALKEAEHYLGLDK 64
```

```
            S I K+I +    + F   + F+ I+LS V+ AE YL  + I  + E L E  H  GLD+
Sbjct: 3    SYIAKRIFAVIPIVLFAIFIMFVFIRLSPVDPAEAYLTAANIHPTEELLAEKRHEFGLDQ 62

Query: 65   PLWKQYWLWFQKALTGDFGYSYVLRLPVLDLVLQRFLATLFLGTSAFLLIVTISTPLGVW 124
            P+  QY     K    DFG+SYV   PV D V  R  ATL L  S+  L V IS PLG
Sbjct: 63   PMAVQYVQTIVKVFQLDFGHSYVTNQPVWDEVTARMPATLQLAVSSIFLAVLISIPLGFL 122

Query: 125  AGLHESARSDHLIRFLSFSSVSMPNFWVAYLLMLLFSAKLNLLPVSGGNDLQSLILPSIT 184
            + +++++  D   R LS+   S+P FW+ YLL+  FS KLNL PV G      L+LP++T
Sbjct: 123  SAIYKNSLIDRFSRLLSYLGASIPQFWLGYLLIFFFSVKLNLFPVEGRGSWAHLVLPTVT 182

Query: 185  LSFSTVGQYIALIRKAISQENRSLNVENARLRGVKERYIVTHHLLRNALPAIMTALSLTW 244
            LS + +  Y  L+R ++ ++ +   V  AR RG+KE+ I+  H+L+ A+  ++T L +
Sbjct: 183  LSLALIAIYTRLLRASVLEQMQESYVLYARTRGIKEKVIMVKHVLKLAISPVITGLGMNV 242

Query: 245  VYLLTGSIIVEEIFSWNGIGRLFVTSLRTSDLPVIQACMLIFGTLFLANNFMTQCFMNWV 304
              LLTG+IIVE++FSW G GR FV ++   D+PVIQ  +L+   LF+  N +       +
Sbjct: 243  GKLLTGTIIVEQVFSWPGFGRYFVDAIFNRDIPVIQCYVLLAACLFIVCNLIVDLVQLAM 302

Query: 305  DPRL 308
            DPR+
Sbjct: 303  DPRI 306
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 63> which encodes the amino acid sequence <SEQ ID 64>. Analysis of this protein sequence reveals the following:

```
     Possible site: 40
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.27   Transmembrane  290 - 306 ( 287 - 313)
   INTEGRAL    Likelihood = -6.37   Transmembrane   12 -  28 (   4 -  33)
   INTEGRAL    Likelihood = -5.89   Transmembrane  105 - 121 ( 100 - 128)
   INTEGRAL    Likelihood = -5.26   Transmembrane  145 - 161 ( 142 - 172)
   INTEGRAL    Likelihood = -2.39   Transmembrane  191 - 207 ( 190 - 208)

----- Final Results -----
               bacterial membrane --- Certainty=0.3909(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 102/324 (31%), Positives = 167/324 (51%), Gaps = 28/324 (8%)

Query: 7    IIKKILSAFLALFFISLLTFILIKLSTVN---SAENYLRLSKISVSPEALKEAEHYLGLD 63
            II KI+     +F +S+LTF+L+K S V+   ++ NY      S++P   K   H+ GLD
Sbjct: 8    IIWKIIRCVTLIFGVSVLTFVLLKQSPVDPVMASVNY----DTSLTPAQYKAIAHHYGLD 63

Query: 64   KPLWKQYWLWFQKALTGDFGYSYVLRLPVLDLVLQRFLATLFLGTSAFLLIVTISTPLGV 123
            KP   QY++W +  + GD G S V R PV D++  R  A+  L   +++L   I   LG
Sbjct: 64   KPALVQYFIWLKNVIQGDLGTSLVYRQPVSDIIRSRAGASFILMGLSWILSGLIGFILGT 123

Query: 124  WAGLHESARSDHLIRFLSFSSVSMPNFWVAYLLMLLFSAKLNLLPVSGGNDL-------- 175
             +  H+    D ++R+ S+  +S+P FW+  + +L+FS +L   P+   + +
Sbjct: 124  LSAFHQGKLLDRVVRWFSYLQISVPTFWIGLIFLLIFSVQLGWFPIGISSPIGTLSQDIT 183

Query: 176  -----QSLILPSITLSFSTVGQYIALIRKAISQENRSLNVENARLRGVKERYIVTHHLLR 230
                 + L+LP  TLS   +       R  +     S  V  AR RG  +  I  HH LR
Sbjct: 184  LADRVKHLMLPVFTLSILGIANVTLHTRTKMMSVLSSEYVLFARARGETQWQIFKHHCLR 243

Query: 231  NALPAIMTALSLTWVY---LLTGSIIVEEIFSWNGIGRLFVTSLRTSDLPVIQACMLIFG 287
            N   AI+ A++L + Y   L  GS++ E++FS+ G+G     +   SD P++ A ++I G
Sbjct: 244  N---AIVPAITLHFSYFGELFGGSVLAEQVFSYPGLGSTLTEAGLKSDTPLLLAIVMI-G 299

Query: 288  TLFL-ANNFMTQCFMNWVDPRLRK 310
            TLF+ A N +     + ++P+LR+
Sbjct: 300  TLFVFAGNLIADILNSIINPQLRR 323
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 23

A DNA sequence (GBSx0020) was identified in *S.agalactiae* <SEQ ID 65> which encodes the amino acid sequence <SEQ ID 66>. This protein is predicted to be nickel transport system (permease). Analysis of this protein sequence reveals the following:

```
Possible site: 14

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -7.64   Transmembrane   57 -  73 (  51 -  80)
   INTEGRAL    Likelihood = -6.85   Transmembrane  173 - 189 ( 169 - 194)
   INTEGRAL    Likelihood = -5.79   Transmembrane   94 - 110 (  86 - 112)
   INTEGRAL    Likelihood = -1.44   Transmembrane  221 - 237 ( 221 - 238)
   INTEGRAL    Likelihood = -1.33   Transmembrane  118 - 134 ( 118 - 134)

----- Final Results -----
               bacterial membrane --- Certainty=0.4057(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB04288 GB:AP001509 nickel transport system (permease)
          [Bacillus halodurans]
 Identities = 103/239 (43%), Positives = 157/239 (65%)

Query: 6   AIFAPILSSFDPQYVDLSQKLLAPNNVHLLGTDQLGRDVLSRLLYGARYSLFLAIIISLL 65
           AI AP ++  DP  V+L+ KLL P+  + LGTDQLGR  LSRLL+GAR SL  A +I +
Sbjct: 29  AILAPWIAPHDPIQVNLALKLLPPSWEYPLGTDQLGRCNLSRLLFGARVSLGFATLIFIS 88

Query: 66  ELTIGMFVGLIVGWYQGKLENLFLWIANIILAFPSFLLSLATVGILGHGLGNLIFAIVFV 125
            L IG+ VG I G+  G ++++ +      ++AFP+ +L L  VG+ G GL  ++ A+V V
Sbjct: 89  SLGIGLLVGAIAGYRGGWIDSVLMRFCEGVMAFPNLVLVLGLVGLFGPGLWQVVLALVMV 148

Query: 126 EWVYYAKLMTNLVKSAKKEPYVINAQIMGLSVWHILRKHIFPFVYQPILVMVLMNIGNII 185
           +WVYYA++  +++ S K++ ++  A+I G S W I+R+HI P V  PI+V+  + +G  I
Sbjct: 149 QWVYYARMFRSMIVSLKEQNFITAARISGSSPWKIIRRHIIPNVLPPIVVIGTLEMGWAI 208

Query: 186 LMISGFSFLGIGVQPNVTEWGMMLHDARGYFRTATWMMLSPGIAIFLTVFSFNTLGDAI 244
           + IS  SFLG+G+QP   EWG M+H+ + + R+   +ML PGI I L V +FN LG+++
Sbjct: 209 MDISALSFLGLGIQPPTPEWGAMIHEGKSFIRSHPELMLYPGIMILLVVMTFNVLGESL 267
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 67> which encodes the amino acid sequence <SEQ ID 68>. Analysis of this protein sequence reveals the following:

```
     Possible site: 39
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.80   Transmembrane  182 - 198 ( 180 - 204)
   INTEGRAL    Likelihood = -7.38   Transmembrane   77 -  93 (  69 -  98)
   INTEGRAL    Likelihood = -7.06   Transmembrane  112 - 128 ( 104 - 132)
   INTEGRAL    Likelihood = -6.16   Transmembrane    8 -  24 (   7 -  31)
   INTEGRAL    Likelihood = -5.10   Transmembrane  239 - 255 ( 235 - 258)

----- Final Results -----
               bacterial membrane --- Certainty=0.4121(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 61/246 (24%), Positives = 127/246 (50%), Gaps = 1/246 (0%)
```

```
Query: 2   LVISAIFAPILSSFDPQYVDLSQKLLAPNNVHLLGTDQLGRDVLSRLLYGARYSLFLAII 61
           L++S +   +     P   + + + LAP+  HL GTD LGRD+  R + G  +SL + ++
Sbjct: 19  LILSILALNLYFYRTPLETNAALRNLAPSLNHLFGTDGLGRDMFVRTIKGLYFSLQVGLL 78

Query: 62  ISLLELTIGMFVGLIVGWYQGKLENLFLWIANIILAFPSFLLSLATVGILGHGLGNLIFA 121
            +L+ + +    G++ G     ++ +  W+ ++ +  P  +  +    ++G G   +I A
Sbjct: 79  GALMGVFLATVFGVLAGLGNSLIDKIIAWLVDLFIGMPHLIFMILISFVVGKGAQGVIIA 138

Query: 122 IVFVEWVYYAKLMTNLVKSAKKEPYVINAQIMGLSVWHILRKHIFPFVYQPILVMVLMNI 181
                W   A+L+ N V   K + +V  ++ MG + ++I+R HI P +   I +  ++
Sbjct: 139 TAVTHWPSLARLIRNEVYDLKNKAFVQLSKSMGKTPYYIVRHHILPLIASQIFIGFILLF 198

Query: 182 GNIILMISGFSFLGIGVQPNVTEWGMMLHDARGYFRTAT-WMMLSPGIAIFLTVFSFNTL 240
            ++IL  +  +FLG G+       G++L +A  +      W+++ PG+ + L V +F+T+
Sbjct: 199 PHVILHEASMTFLGFGLSAEQPSVGIILSEAAKHISLGNWWLVIFPGLYLILVVNAFDTI 258

Query: 241 GDAIDK 246
           G+++ K
Sbjct: 259 GESLKK 264
```

A related GBS gene <SEQ ID 8473> and protein <SEQ ID 8474> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 0
McG: Discrim Score:      7.56
GvH: Signal Score (-7.5): -1.15
     Possible site: 14
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 5 value:  -7.64 threshold:  0.0
   INTEGRAL    Likelihood = -7.64   Transmembrane   57 -  73 (  51 -  80)
   INTEGRAL    Likelihood = -6.85   Transmembrane  173 - 189 ( 169 - 194)
   INTEGRAL    Likelihood = -5.79   Transmembrane   94 - 110 (  86 - 112)
   INTEGRAL    Likelihood = -1.44   Transmembrane  221 - 237 ( 221 - 238)
   INTEGRAL    Likelihood = -1.33   Transmembrane  118 - 134 ( 118 - 134)
   PERIPHERAL  Likelihood =  4.72     145
 modified ALOM score:   2.03

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4057(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF02082(292 - 1053 of 1365)
EGAD|89511|HP0300(23 - 283 of 285) dipeptide ABC transporter, permease protein (dppC)
{Helicobacter pylori} OMNI|HP0300 dipeptide ABC transporter, permease protein (dppC)
GP|2313398|gb|AAD07369.1||AE000548 dipeptide ABC transporter, permease protein (dppC)
{Helicobacter pylori 26695} PIR|D64557|D64557 dipeptide ABC transporter, permease protein -
Helicobacter pylori (strain 26695)
%Match = 20.5
%Identity = 43.4  %Similarity = 63.3
Matches = 111  Mismatches = 92  Conservative Sub.s = 51

30        60        90        120       150       180       210       240
P*KCLTCDNDST*LDLGLLINRINYC*RNFFMEWNRTFICDQSKNFRSSSNTSLYANFWNLIFS**FYDTVFYELG*SSV

                                                                      MESFR

270       300       330       360                 402       432       462
TKVKGEIISKRIYFSSSLLVLLVISAIFAPILSSFDPQYVDLSQKLLAP------NNVHLLGTDQLGRDVLSRLLYGARY
                  :::|||| |||||:|:  ||   :   :|| |      |  :||||| ||||:||||:||||
EFIQQFKKNKAAVVGAWIVLLLVICAIFAPLLAPHDPYVQNAQDRLLKPIWEHGGNAKYLLGTDDLGRDILSRLIYGARI
             20        30        40        50        60        70        80
```

```
492       522       552       582       612       642       672       702
SLFLAIIISLLELTIGMFVGLIVGWYQGKLENLFLWIANIILAFPSFLLSLATVGILGHGLGNLIFAIVFVEWVYYAKLM
|| : |:   : :  |  :||| |:: || : : : | :|::|:|| || :  | :||  | | ::|| ||    :|:|:
SLTIGIVSMGIAVFFGTILGLIAGYFGGKTDAIIMRIMDIMFALPSILLIVIVVAVLGPSLTNAMLAIGFVGIPGFARLV
            100       110       120       130       140       150       160

732       762       792       822       852       882       912       942
TNLVKSAKKEPYVINAQIMGLSVWHILRKHIFPFVYQPILVMVLMNIGNIILMISGFSFLGIGVQPNVTEWGMMLHDARG
 : |   |:: ||| ::| | |   :: | |||    |::|   |   : :|  : :||||:| ||   ||| || ::
RSSVLGEKEKEYVIASKINGSSHLRLMCKVIFPNCIIPLIVQTTMGFASTVLEAAALSFLGLGAQPPKPEWGAMLMNSMQ
            180       190       200       210       220       230       240

972      1002      1032      1059      1089      1119      1149
YFRTATWMMLSPGIAIFLTVFSFNTLGDAI-DKKDWKRQWNS*K*ENCHYR*ERSLY*EILVVK*IWENR*LLLVRVV
|  || ||:: ||: ||||| ||| :|| | |  | ||
YIATAPWMLVFPGVMIFLTVMSFNLVGDGIMDALDPKRTS
            260       270       280
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 24

A DNA sequence (GBSx0021) was identified in *S.agalactiae* <SEQ ID 69> which encodes the amino acid sequence <SEQ ID 70>. This protein is predicted to be peptide ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

```
Possible site: 60

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.32   Transmembrane  161 - 177 ( 161 - 177)

----- Final Results -----
               bacterial membrane --- Certainty=0.1128(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10027> which encodes amino acid sequence <SEQ ID 10028> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF73561 GB:AE002315 peptide ABC transporter, ATP-binding
           protein [Chlamydia muridarum]
 Identities = 86/253 (33%), Positives = 154/253 (59%), Gaps = 2/253 (0%)

Query: 1   METTMEQLEIRKLSLQIGEVPVLRDFSCKIDMGESLTIIGESGSGKTLLAKLLVGHIPQG 60
           M  T+ ++E   ++++     ++   S  I   +SL ++GE+GSGKT ++K ++G +P
Sbjct: 1   MSKTLLKIENLVVAIKESNQRLVNHLSLTIKQRQSLALVGENGSGKTTVSKAILGFLPDN 60

Query: 61  MTVR-GNIFFKGVDLGKLTVKQWQKLRGRDIAYLVQNPMSMFNPFQKIEAHILETILSHE 119
             ++ G IF+ G D+ +L+ K++Q +RG+ I+ + QN M    P  ++   I+ET+  H
Sbjct: 61  CCIQSGKIFYSGTDITRLSRKEFQSIRGKKISTIFQNAMGTLTPSMRVGTQIIETLRHHF 120

Query: 120 KCSKRVALSKALEWMKRLNLDDAISLLKKYPFELSGGMLQRIMLATILSLDPQVIILDEP 179
             SK  A +KA E +  ++++     L+ YPFELSGGM QR+ +A  L+ +P++II DEP
Sbjct: 121 VMSKEEAFAKARELLVSVHIESPDRCLQLYPFELSGGMCQRVSIAIALATNPELIIADEP 180

Query: 180 TSAVDCHNCSTISAILQEL-QNNGKTLITVTHDYQLARDLGGQLLVISEGEVVEQGQTQA 238
           ++A+D  + + +  +L+++ QNN   L+ +TH+  L  +L  ++ +I  GE+VEQG
Sbjct: 181 STALDSISQAQVLRVLKQIHQNNNTALLLITHNLALVSELCEEMAIIHHGEIVEQGPVHE 240

Query: 239 ILSNPQHNYTKAL 251
```

```
          +L +P H YT+ L
Sbjct: 241 LLRSPSHPYTQKL 253
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 71> which encodes the amino acid sequence <SEQ ID 72>. Analysis of this protein sequence reveals the following:

```
      Possible site: 55

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.50   Transmembrane  168 - 184 ( 167 - 184)
   INTEGRAL    Likelihood = -1.70   Transmembrane  211 - 227 ( 211 - 227)

----- Final Results -----
               bacterial membrane --- Certainty=0.1999(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 87/232 (37%), Positives = 138/232 (58%), Gaps = 3/232 (1%)

Query: 23  LRDFSCKIDMGESLTIIGESGSGKTLLAKLLVGHIPQ-GMTVRGNIFFKGVDLGKL-TVK 80
           +R+ S ++  GE L  +GESGSGK++L K   G +   G    G+I ++G +L  L T K
Sbjct: 28  IRNVSLELVEGEVLAFVGESGSGKSVLTKTFTGMLESNGRIANGSIVYRGQELTDLKTNK 87

Query: 81  QWQKLRGRDIAYLVQNPMSMFNPFQKIEAHILETILSHEKCSKRVALSKALEWMKRLNLD 140
           +W K+RG  IA + Q+PM+  +P + I + I E I+ H+K S   A   AL++M ++ +
Sbjct: 88  EWAKIRGSKIATIFQDPMTSLSPIKTIGSQITEVIIKHQKVSHAKAKEMALDYMNKVGIP 147

Query: 141 DAISLLKKYPFELSGGMLQRIMLATILSLDPQVIILDEPTSAVDCHNCSTISAILQELQN 200
           +A    + YPFE SGGM QRI++A  L+  P ++I DEPT+A+D    + I  +L+ LQ
Sbjct: 148 NAKKRFEDYPFEYSGGMRQRIVIAIALACRPDILICDEPTTALDVTIQAQIVELLKSLQR 207

Query: 201 NGK-TLITVTHDYQLARDLGGQLLVISEGEVVEQGQTQAILSNPQHNYTKAL 251
               T+I +THD  +   + ++ V+  GE+VE G  + I  +P+H YT +L
Sbjct: 208 EYHFTIIFITHDLGVVASIADKVAVMYAGEIVEFGTVEEIFYDPRHPYTWSL 259
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 25

A DNA sequence (GBSx0022) was identified in *S.agalactiae* <SEQ ID 73> which encodes the amino acid sequence <SEQ ID 74>. This protein is predicted to be peptide ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

```
Possible site: 50

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10025> which encodes amino acid sequence <SEQ ID 10026> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB05797 GB:AP001514 oligopeptide ABC transporter (ATP-binding
          protein) [Bacillus halodurans]
 Identities = 82/199 (41%), Positives = 130/199 (65%), Gaps = 2/199 (1%)
```

```
Query: 19  RQEVLKDCHFHLKRGEIIGIMGKSGSGKSSLARLIIGLDSPTCGSIYFQG-KIYTPKDGK 77
           +Q++L    F  + GE +GI+G+SGSGKS+L RL++G++ P  G IYF+G K+
Sbjct: 21  KQKILNHISFECRHGECLGIIGESGSGKSTLGRLLLGIEKPDRGHIYFEGNKVEERSVRS 80

Query: 78  AQIILVFQDALSSVNPYFSIEEILNEAFYGKKTT-FELCQILEAVGLDGTYLKYKARQLS 136
             I  VFQD  SS+NP+F++E  + E   GKK    ++  +L+ VGL  +Y K    +LS
Sbjct: 81  GNISAVFQDYTSSINPFFTVETAIMEPLKGKKAAKSKVDYLLKQVGLHPSYKKKYPHELS 140

Query: 137 GGQLQRVCIARALLLKPKIIIFDESLSGLDPVTQIKMLRLLQKIKRRYELSFIMISHDPK 196
           GG++QRVCIARA+  +PK I+ DE++S LD   Q ++L LL ++KR Y++S++ I+HD +
Sbjct: 141 GGEVQRVCIARAISTEPKCIVLDEAISSLDVSIQTQVLDLLIELKRIYQMSYLFITHDIQ 200

Query: 197 ICQAICNRVFLIKNGYLVE 215
               IC+R+ + ++G + E
Sbjct: 201 AAAYICDRIMIFRHGQIEE 219
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 75> which encodes the amino acid sequence <SEQ ID 76>. Analysis of this protein sequence reveals the following:

```
Possible site: 60

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3195(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 91/238 (38%), Positives = 137/238 (57%), Gaps = 21/238 (8%)

Query: 1   MKEIFLMLVCNHVGKTFGRQ----EVLKDCHFHLKRGEIIGIMGKSGSGKSSLARLIIGL 56
           M E  + L  +H+  TF ++    E +KD   H+ +G+I GI+G SG+GKS+L R+I  L
Sbjct: 1   MNEAIIQL--DHIDITFRQKKRVIEAVKDVTVHINQGDIYGIVGYSGAGKSTLVRVINLL 58

Query: 57  DSPTCGSI-------YFQGKIYTPKDGKAQ----IILVFQ--DALSSVNPYFSIEEILNE 103
            +PT G I       + QGKI    D   Q    I ++FQ  + ++      ++   L
Sbjct: 59  QAPTNGKITVDGDVTFDQGKIQLSADALRQKRRDIGMIFQHFNLMAQKTAKENVAFALRH 118

Query: 104 AFYGK-KTTFELCQILEAVGLDGTYLKYKARQLSGGQLQRVCIARALLLKPKIIIFDESL 162
           +   K +   ++ ++LE VGL      Y A QLSGGQ QRV IARAL   PKI+I DE+
Sbjct: 119 SSLSKTEKEHKVIELLELVGLSERADNYPA-QLSGGQKQRVAIARALANDPKILISDEAT 177

Query: 163 SGLDPVTQIKMLRLLQKIKRRYELSFIMISHDPKICQAICNRVFLIKNGYLVEDNEFL 220
           S LDP T  ++L LLQ++ R+  L+ +MI+H+ +I + ICNRV +++NG L+E+   L
Sbjct: 178 SALDPKTTKQILALLQELNRKLGLTIVMITHEMQIVKDICNRVAVMQNGVLIEEGSVL 235
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 26

A DNA sequence (GBSx0023) was identified in *S.agalactiae* <SEQ ID 77> which encodes the amino acid sequence <SEQ ID 78>. This protein is predicted to be UMP kinase (pyrH). Analysis of this protein sequence reveals the following:

```
Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1935(Affirmative) < succ>
```

```
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB13524 GB:Z99112 uridylate kinase [Bacillus subtilis]
 Identities = 143/238 (60%), Positives = 193/238 (81%)

Query: 2   EPKYQRILIKLSGEALAGDKGVGIDIPTVQSIAKEIAEVHNSGVQIALVIGGGNLWRGEP 61
           +PKY+RI++KLSGEALAG++G GI+    +QSIAK++ E+     V++A+V+GGGN    +
Sbjct: 3   KPKYKRIVLKLSGEALAGEQGNGINPTVIQSIAKQVKEIAELEVEVAVVVGGGNYGAEKT 62

Query: 62  AAEAGMDRVQADYTGMLGTVMNALVMADSLQQYGVDTRVQTAIPMQTVAEPYVRGRALRH 121
            ++ GMDR  ADY GML TVMN+L + DSL+  G+ +RVQT+I M+ VAEPY+R +A+RH
Sbjct: 63  GSDLGMDRATADYMGMLATVMNSLALQDSLETLGIQSRVQTSIEMRQVAEPYIRRKAIRH 122

Query: 122 LEKNRIVVFGAGIGSPYFSTDTTAALRAAEIEAEAILMAKNGVDGVYNADPKKDANAVKF 181
           LEK R+V+F AG G+PYFSTDTTAALRAAEIEA+ ILMAKN VDGVYNADP+KD +AVK+
Sbjct: 123 LEKKRVVIFAAGTGNPYFSTDTTAALRAAEIEADVILMAKNNVDGVYNADPRKDESAVKY 182

Query: 182 DELTHVEVIKRGLKIMDATASTISMDNDIDLVVFNMNETGNIKRVVLGEQIGTTVSNK 239
           + L++++V+K GL++MD+TAS++ MDNDI L+VF++ E GNIKR V+GE IGT V  K
Sbjct: 183 ESLSYLDVLKDGLEVMDSTASSLCMDNDIPLIVFSIMEEGNIKRAVIGESIGTIVRGK 240
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 79> which encodes the amino acid sequence <SEQ ID 80>. Analysis of this protein sequence reveals the following:

```
Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1955(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 224/242 (92%), Positives = 233/242 (95%)

Query: 1   MEPKYQRILIKLSGEALAGDKGVGIDIPTVQSIAKEIAEVHNSGVQIALVIGGGNLWRGE 60
           +EPKYQRILIKLSGEALAG+KGVGIDIPTVQ+IAKEIAEVH SGVQIALVIGGGNLWRGE
Sbjct: 1   VEPKYQRILIKLSGEALAGEKGVGIDIPTVQAIAKEIAEVHVSGVQIALVIGGGNLWRGE 60

Query: 61  PAAEAGMDRVQADYTGMLGTVMNALVMADSLQQYGVDTRVQTAIPMQTVAEPYVRGRALR 120
           PAA+AGMDRVQADYTGMLGTVMNALVMADSLQ YGVDTRVQTAIPMQ VAEPY+RGRALR
Sbjct: 61  PAADAGMDRVQADYTGMLGTVMNALVMADSLQHYGVDTRVQTAIPMQNVAEPYIRGRALR 120

Query: 121 HLEKNRIVVFGAGIGSPYFSTDTTAALRAAEIEAEAILMAKNGVDGVYNADPKKDANAVK 180
           HLEKNRIVVFGAGIGSPYFSTDTTAALRAAEIEA+AILMAKNGVDGVYNADPKKDANAVK
Sbjct: 121 HLEKNRIVVFGAGIGSPYFSTDTTAALRAAEIEADAILMAKNGVDGVYNADPKKDANAVK 180

Query: 181 FDELTHVEVIKRGLKIMDATASTISMDNDIDLVVFNMNETGNIKRVVLGEQIGTTVSNKA 240
           FDELTH EVIKRGLKIMDATAST+SMDNDIDLVVFNMNE GNI+RVV GE IGTTVSNK
Sbjct: 181 FDELTHGEVIKRGLKIMDATASTLSMDNDIDLVVFNMNEAGNIQRVVFGEHIGTTVSNKV 240

Query: 241 SE 242
            +
Sbjct: 241 CD 242
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 27

A DNA sequence (GBSx0024) was identified in *S.agalactiae* <SEQ ID 81> which encodes the amino acid sequence <SEQ ID 82>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3712(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 28

A DNA sequence (GBSx0025) was identified in *S.agalactiae* <SEQ ID 83> which encodes the amino acid sequence <SEQ ID 84>. This protein is predicted to be ribosome recycling factor (frr). Analysis of this protein sequence reveals the following:

```
Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3522(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB06143 GB:AP001515 ribosome recycling factor [Bacillus halodurans]
 Identities = 112/185 (60%), Positives = 149/185 (80%)

Query: 1   MTKEIVTKAQERFEQSHQSLSREFAGIRAGRANASLLDRIQVEYYGAPTPLNQLASITVP 60
           M+KE++  A++R  ++ ++L RE A +RAGRAN ++LDRI VEYYGA TPLNQLA+I+VP
Sbjct: 1   MSKEVLNDAEQRMTKATEALGRELAKLRAGRANPAMLDRITVEYYGAETPLNQLATISVP 60

Query: 61  EARVLLISPFDKSSIKDIERAINESDLGINPANDGSVIRLVIPALTEETRRDLAKEVKKV 120
           EAR+L+I PFDKSSI DIERAI +SDLG+ P+NDG+VIR+ IP LTEE RRDL K VKK
Sbjct: 61  EARLLVIQPFDKSSISDIERAIQKSDLGLTPSNDGTVIRITIPPLTEERRRDLTKLVKKS 120

Query: 121 GENAKIAIRNIRRDAMDEAKKQEKNKEITEDDLKSLEKDIQKATDDAVKHIDEMTANKEK 180
            E AK+A+RNIRRDA D+ KK++K+ E+TEDDL+ + +D+QK TD  ++ ID+    KEK
Sbjct: 121 AEEAKVAVRNIRRDANDDLKKRQKDGELTEDDLRRVTEDVQKLTDKYIEQIDQKAEAKEK 180

Query: 181 ELLEV 185
           E++EV
Sbjct: 181 EIMEV 185
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 85> which encodes the amino acid sequence <SEQ ID 86>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4462(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 160/185 (86%), Positives = 171/185 (91%)

Query: 1   MTKEIVTKAQERFEQSHQSLSREFAGIRAGRANASLLDRIQVEYYGAPTPLNQLASITVP 60
           M   I+  A+ERF QSHQSLSRE+A IRAGRANASLLDRIQV+YYGAPTPLNQLASITVP
Sbjct: 1   MANAIIETAKERFAQSHQSLSREYASIRAGRANASLLDRIQVDYYGAPTPLNQLASITVP 60

Query: 61  EARVLLISPFDKSSIKDIERAINESDLGINPANDGSVIRLVIPALTEETRRDLAKEVKKV 120
           EARVLLISPFDKSSIKDIERA+N SDLGI PANDGSVIRLVIPALTEETR++LAKEVKKV
Sbjct: 61  EARVLLISPFDKSSIKDIERALNASDLGITPANDGSVIRLVIPALTEETRKELAKEVKKV 120

Query: 121 GENAKIAIRNIRRDAMDEAKKQEKNKEITEDDLKSLEKDIQKATDDAVKHIDEMTANKEK 180
           GENAKIAIRNIRRDAMD+AKKQEK KEITED+LK+LEKDIQKATDDA+K ID MTA KEK
Sbjct: 121 GENAKIAIRNIRRDAMDDAKKQEKAKEITEDELKTLEKDIQKATDDAIKEIDRMTAEKEK 180

Query: 181 ELLEV 185
           ELL V
Sbjct: 181 ELLSV 185
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 29

A DNA sequence (GBSx0026) was identified in *S.agalactiae* <SEQ ID 87> which encodes the amino acid sequence <SEQ ID 88>. Analysis of this protein sequence reveals the following:

```
Possible site: 44

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1356(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10023> which encodes amino acid sequence <SEQ ID 10024> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB12943 GB:Z99109 yitL [Bacillus subtilis]
 Identities = 107/269 (39%), Positives = 155/269 (56%), Gaps = 6/269 (2%)

Query: 42  LVTDENKDF-YFIQKDGFTFALSKSEGEHHIGEM--VKGFAYTDMQQKARLTTKETFATR 98
           L  D   DF YF+     T  L  SE    I +   V+ F Y D Q++   T K    +
Sbjct: 25  LSIDHQTDFGYFLTDGEDTILLHNSEMTEDIEDRDEVEVFIYVDQQERLAATMKIPIISA 84

Query: 99  DHYGWGTVTEVRKDLGVFLDTGLPDKQVVVSLDVLPELKELWPKKGDRLYVCLDVDKKDR 158
           D YGW  V +  +D+GVF+D GL  K  +V+ + LP  +++WP+KGD+LY  L V  + R
Sbjct: 85  DEYGWVEVVDKVEDMGVFVDVGL-SKDALVATEHLPPYEDVWPQKGDKLYCMLKVTNRGR 143

Query: 159 LWALPADPEVFQRMATPAYNNMQNQNWPAIVYRLKLSGTFVYLPENNMLGFIHPSERYSE 218
           ++A PA  ++   + T A  ++ N+     VYRL  SG+FV + ++ +  FIHPSER  E
Sbjct: 144 MFAKPAPEDIISELFTDASEDLMNKELTGTVYRLIASGSFV-ITDDGIRCFIHPSERKEE 202

Query: 219 PRLGQVLDARVIGFREVDRTLNLSLKPRSFEMLENDAQMILTYLESNGGFMTLNDKSSPE 278
           PRLG  +  RVI  +E D ++NLSL PR  + +  DA+ ILTY+    G M  +DKS P+
```

```
Sbjct: 203 PRLGSRVTGRVIQVKE-DGSVNLSLLPRKQDAMSVDAECILTYMRMRNGAMPYSDKSQPD 261

Query: 279 EIKATFGISKGQFKKALGGLMKAKKIKQD 307
           +I+  F +SK  FK+ALG LMK  K+ Q+
Sbjct: 262 DIRERFNMSKAAFKRALGHLMKNGKVYQE 290
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 89> which encodes the amino acid sequence <SEQ ID 90>. Analysis of this protein sequence reveals the following:

```
Possible site: 51

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0811(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 235/284 (82%), Positives = 265/284 (92%)

Query: 31  MNTLLATVITGLVTDENKDFYFIQKDGFTFALSKSEGEHHIGEMVKGFAYTDMQQKARLT 90
           MN LLATVITGL+ +EN + YFI K+GFTF LSK+EGE  IG+MV GFAYTD++QKARLT
Sbjct: 1   MNDLLATVITGLIKEENANDYFIHKEGFTFTLSKAEGERQIGDMVTGFAYTDIEQKARLT 60

Query: 91  TKETFATRDHYGWGTVTEVRKDLGVFLDTGLPDKQVVVSLDVLPELKELWPKKGDRLYVC 150
           TKE  +TR  YGWG VTEVR+DLGVF+DTG+P+K++VVSLDVLPE+KELWPKKGD+LY+
Sbjct: 61  TKEIRSTRTSYGWGEVTEVRRDLGVFVDTGIPNKEIVVSLDVLPEMKELWPKKGDKLYIR 120

Query: 151 LDVDKKDRLWALPADPEVFQRMATPAYNNMQNQNWPAIVYRLKLSGTFVYLPENNMLGFI 210
           LDVDKKDR+W LPA+PEVFQ+MA+PAYNNMQNQ+WPAIVYRLKL+GTFVYLPENNMLGFI
Sbjct: 121 LDVDKKDRIWGLPAEPEVFQKMASPAYNNMQNQHWPAIVYRLKLTGTFVYLPENNMLGFI 180

Query: 211 HPSERYSEPRLGQVLDARVIGFREVDRTLNLSLKPRSFEMLENDAQMILTYLESNGGFMT 270
           H SERY+EPRLGQVLDARVIGFREVDRTLNLSLKPRSFEMLENDAQMI+TYLE+NGGFMT
Sbjct: 181 HSSERYAEPRLGQVLDARVIGFREVDRTLNLSLKPRSFEMLENDAQMIVTYLEANGGFMT 240

Query: 271 LNDKSSPEEIKATFGISKGQFKKALGGLMKAKKIKQDQLGTELL 314
           LNDKSSPEEIKA+FGISKGQFKKALGGLMKAK+IKQD  GTEL+
Sbjct: 241 LNDKSSPEEIKASFGISKGQFKKALGGLMKAKRIKQDATGTELI 284
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 30

A DNA sequence (GBSx0028) was identified in *S.agalactiae* <SEQ ID 91> which encodes the amino acid sequence <SEQ ID 92>. This protein is predicted to be peptide methionine sulfoxide reductase (msrA). Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0866(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10021> which encodes amino acid sequence <SEQ ID 10022> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB05167 GB:AP001512 peptide methionine sulfoxide reductase
            [Bacillus halodurans]
 Identities = 102/173 (58%), Positives = 126/173 (71%), Gaps = 2/173 (1%)

Query: 14  ENDMERAIFAGGCFWCMVQPFEELDGIESVLSGYTGGHVENPTYKEVCSKTTGHTEAVEI 73
           E+    A FAGGCFWCMV PFEE  GI  V+SGYTGGH ENPTYKEVCS+TTGH EAV+I
Sbjct: 3   ESKWALATFAGGCFWCMVSPFEEEPGIHQVVSGYTGGHTENPTYKEVCSETTGHYEAVQI 62

Query: 74  IFNPEKISYADLVELYWAQTDPTDAFGQFEDRGDNYRPVIFYENEEQRQIAQKSKDKLQA 133
            F+PE   Y  L+E+YW Q DPTD  GQF DRGD+YR  IFY +E+Q+Q A  SK KL+
Sbjct: 63  SFDPEVFPYEKLLEIYWTQIDPTDPGGQFHDRGDSYRTAIFYHDEQQKQAADASKQKLEE 122

Query: 134 SGRFDRPIVTSIEPADTFYPAEDYHQAFYRTNPARYAL--SSARRHAFLEENW 184
           SG+F+ PIVT I PA  FYPAE+YHQ +++ NP  Y +    + R AF++++W
Sbjct: 123 SGKFNAPIVTRILPAKPFYPAEEYHQKYHKKNPFHYKMYRHGSGREAFIKQHW 175
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 93> which encodes the amino acid sequence <SEQ ID 94>. Analysis of this protein sequence reveals the following:

```
     Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0084(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 89-91
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB05167 GB:AP001512 peptide methionine sulfoxide reductase
            [Bacillus halodurans]
 Identities = 98/168 (58%), Positives = 125/168 (74%), Gaps = 4/168 (2%)

Query: 4   AIFAGGCFWCMVQPFEEQAGILSVRSGYTGGHLPNPSYEQVCAKTTGHTEAVEIIFDPKQ 63
           A FAGGCFWCMV PFEE+ GI  V SGYTGGH  NP+Y++VC++TTGH EAV+I FDP+
Sbjct: 9   ATFAGGCFWCMVSPFEEEPGIHQVVSGYTGGHTENPTYKEVCSETTGHYEAVQISFDPEV 68

Query: 64  IAYKDLVELYWTQTDPTDAFGQFEDRGDNYRPVIYYTTERQKEIAEQSKANLQASGRFDQ 123
             Y+ L+E+YWTQ DPTD  GQF DRGD+YR  I+Y  E+QK+ A+ SK  L+ SG+F+
Sbjct: 69  FPYEKLLEIYWTQIDPTDPGGQFHDRGDSYRTAIFYHDEQQKQAADASKQKLEESGKFNA 128

Query: 124 PIVTTIEPAEPFYLAEDYHQGFYKKNP---KRYAQSSAIRHQFLEENW 168
           PIVT I PA+PFY AE+YHQ ++KKNP   K Y   S  R  F++++W
Sbjct: 129 PIVTRILPAKPFYPAEEYHQKYHKKNPFHYKMYRHGSG-REAFIKQHW 175
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 130/168 (77%), Positives = 148/168 (87%)

Query: 17  MERAIFAGGCFWCMVQPFEELDGIESVLSGYTGGHVENPTYKEVCSKTTGHTEAVEIIFN 76
           MERAIFAGGCFWCMVQPFEE  GI SV SGYTGGH+ NP+Y++VC+KTTGHTEAVEIIF+
Sbjct: 1   MERAIFAGGCFWCMVQPFEEQAGILSVRSGYTGGHLPNPSYEQVCAKTTGHTEAVEIIFD 60

Query: 77  PEKISYADLVELYWAQTDPTDAFGQFEDRGDNYRPVIFYENEEQRQIAQKSKDKLQASGR 136
           P++I+Y DLVELYW QTDPTDAFGQFEDRGDNYRPVI+Y  E Q++IA++SK  LQASGR
Sbjct: 61  PKQIAYKDLVELYWTQTDPTDAFGQFEDRGDNYRPVIYYTTERQKEIAEQSKANLQASGR 120

Query: 137 FDRPIVTSIEPADTFYPAEDYHQAFYRTNPARYALSSARRHAFLEENW 184
           FD+PIVT+IEPA+ FY AEDYHQ FY+ NP RYA SSA RH FLEENW
Sbjct: 121 FDQPIVTTIEPAEPFYLAEDYHQGFYKKNPKRYAQSSAIRHQFLEENW 168
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 31

A DNA sequence (GBSx0029) was identified in *S.agalactiae* <SEQ ID 95> which encodes the amino acid sequence <SEQ ID 96>. Analysis of this protein sequence reveals the following:

```
Possible site: 55

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2727(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB13859 GB:Z99114 yozE [Bacillus subtilis]
 Identities = 24/66 (36%), Positives = 42/66 (63%)

Query: 3  KSFYSWLMTQRNPKSNEPVAILADYAFDETTFPKHSSDFETVSRYLEDEASFSFNLTDFD 62
          KSFY +L+  R+PK  + ++  A+ A+++ +FPK S+D+  +S YLE  A +   +  FD
Sbjct: 2  KSFYHYLLKYRHPKPKDSISEFANQAYEDHSFPKTSTDYHEISSYLELNADYLHTMATFD 61

Query: 63 DIWEDY 68
          + W+ Y
Sbjct: 62 EAWDQY 67
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 97> which encodes the amino acid sequence <SEQ ID 98>. Analysis of this protein sequence reveals the following:

```
Possible site: 57

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2571(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 59/71 (83%), Positives = 65/71 (91%)

Query: 1  MRKSFYSWLMTQRNPKSNEPVAILADYAFDETTFPKHSSDFETVSRYLEDEASFSFNLTD 60
          MRKSFYSWLMTQRNPKSNEPVAILAD  FD+TTFPKH++DFE +SRYLED+ASFSFNL
Sbjct: 3  MRKSFYSWLMTQRNPKSNEPVAILADLVFDDTTFPKHTNDFELISRYLEDQASFSFNLGQ 62

Query: 61 FDDIWEDYLNH 71
          FD+IWEDYL H
Sbjct: 63 FDEIWEDYLAH 73
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 32

A DNA sequence (GBSx0030) was identified in *S.agalactiae* <SEQ ID 99> which encodes the amino acid sequence <SEQ ID 100>. This protein is predicted to be antigen, 67 kDa (myosin-crossreactive). Analysis of this protein sequence reveals the following:

```
Possible site: 14

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.57   Transmembrane   28 -  44 (  26 -  45)

----- Final Results -----
               bacterial membrane --- Certainty=0.2826(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 101> which encodes the amino acid sequence <SEQ ID 102>. Analysis of this protein sequence reveals the following:

```
Possible site: 26

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -4.62   Transmembrane   40 -  56 (  38 -  57)

----- Final Results -----
               bacterial membrane --- Certainty=0.2848(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9109> which encodes the amino acid sequence <SEQ ID 9110>. Analysis of this protein sequence reveals the following:

```
     Possible cleavage site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial membrane --- Certainty= 0.285(Affirmative) < succ>
                bacterial outside --- Certainty= 0.000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 477/590 (80%), Positives = 542/590 (91%)

Query: 3   MRYTNGNFEAFARPRKPEGVDKKSAYIVGSGLAGLAAAVFLIRDGQMDGQRIHIFEELPL 62
           M YT+GN+EAFA PRKPEGVD+KSAYIVG+GLAGLAAAVFLIRDG M G+RIH+FEELPL
Sbjct: 15  MYYTSGNYEAFATPRKPEGVDQKSAYIVGTGLAGLAAAVFLIRDGHMAGERIHLFEELPL 74

Query: 63  SGGSLDGVKRPDIGFVTRGGREMENHFECMWDMYRSIPSLEVPDASYLDEFYWLDKDDPN 122
           +GGSLDG+++P +GFVTRGGREMENHFECMWDMYRSIPSLE+P ASYLDEFYWLDKDDPN
Sbjct: 75  AGGSLDGIEKPHLGFVTRGGREMENHFECMWDMYRSIPSLEIPGASYLDEFYWLDKDDPN 134

Query: 123 SSNCRLIHKQGNRLESDGDFTLGTHSKELVKLVMETEESLGAKTIEEVFSKEFFESNFWT 182
           SSNCRLIHK+GNR++ DG +TLG  SKEL+ L+M+TEESLG +TIEE FS++FF+SNFW
Sbjct: 135 SSNCRLIHKRGNRVDDDGQYTLGKQSKELIHLIMKTEESLGDQTIEEFFSEDFFKSNFWV 194

Query: 183 YWGTMFAFEKWHSAIEMRRYAMRFIHHIGGLPDFTSLKFNKYNQYDSMVKPIISYLESHN 242
           YW TMFAFEKWHSA+EMRRYAMRFIHHI GLPDFTSLKFNKYNQYDSMVKPII+YLESH+
Sbjct: 195 YWATMFAFEKWHSAVEMRRYAMRFIHHIDGLPDFTSLKFNKYNQYDSMVKPIIAYLESHD 254

Query: 243 VDVQFDSKVTNISVDFKNGQKLAKAIHLTVGGEAKTIDLTPNDFVFVTNGSITESTNYGS 302
           VD+QFD+KVT+I V+   G+K+AK IH+TV GEAK I+LTP+D VFVTNGSITES+ YGS
Sbjct: 255 VDIQFDTKVTDIQVEQTAGKKVAKTIHMTVSGEAKAIELTPDDLVFVTNGSITESSTYGS 314

Query: 303 HDTVAKPNTDLGGSWNLWENLAAQSDEFGHPKVFYKDIPKESWFVSATATIKDPAIEPYI 362
           H  VAKP   LGGSWNLWENLAAQSD+FGHPKVFY+D+P ESWFVSATATIK PAIEPYI
Sbjct: 315 HHEVAKPTKALGGSWNLWENLAAQSDDFGHPKVFYQDLPAESWFVSATATIKHPAIEPYI 374

Query: 363 ERLTHRDLHDGKVNTGGIVTVTDSNWMMSFAIHRQPHFKEQKENETIVWIYGLYSNVEGN 422
           ERLTHRDLHDGKVNTGGI+T+TDSNWMMSFAIHRQPHFKEQKENET VWIYGLYSN EGN
Sbjct: 375 ERLTHRDLHDGKVNTGGIITITDSNWMMSFAIHRQPHFKEQKENETTVWIYGLYSNSEGN 434
```

```
Query: 423 YIKKPIEECTGREITEEWLYHLGVPEMKIHDLSDKQYVSTVPVYMPYITSYFMPRVKGDR 482
           Y+ K IEECTG+EITEEWLYHLGVP  KI DL+ + Y++TVPVYMPYITSYFMPRVKGDR
Sbjct: 435 YVHKKIEECTGQEITEEWLYHLGVPVDKIKDLASQDYINTVPVYMPYITSYFMPRVKGDR 494

Query: 483 PDVIPQGSVNLAFIGNFAESPSRDTVFTTEYSIRTAMEAVYTFLNIERGVPEVFNSAFDI 542
           P VIP GSVNLAFIGNFAESPSRDTVFTTEYSIRTAMEAVY+FLN+ERG+PEVFNSA+DI
Sbjct: 495 PKVIPDGSVNLAFIGNFAESPSRDTVFTTEYSIRTAMEAVYSFLNVERGIPEVFNSAYDI 554

Query: 543 RVLLQSLYYLNDKKSVEDMDLPIPALMRKVGMKKIRGTYLEELLREAHLL 592
           R LL++ YYLNDKK+++DMDLPIPAL+ K+G KKI+ T++EELL++A+L+
Sbjct: 555 RELLKAFYYLNDKKAIKDMDLPIPALIEKIGHKKIKDTFIEELLKDANLM 604
```

A related GBS gene <SEQ ID 8475> and protein <SEQ ID 8476> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:    -19.82
GvH: Signal Score (-7.5): -1.16
     Possible site: 14
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -4.57 threshold:  0.0
   INTEGRAL    Likelihood = -4.57   Transmembrane   26 -  42 (  26 -  45)
   PERIPHERAL  Likelihood =  6.79     378
 modified ALOM score:   1.41

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.2826(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear)
```

SEQ ID 8476 (GBS90) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 18 (lane 6; MW 68.5kDa).

The GBS90-His fusion product was purified (Figure 194, lane 11) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 256A), FACS (Figure 256B), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 33

A DNA sequence (GBSx0031) was identified in *S.agalactiae* <SEQ ID 103> which encodes the amino acid sequence <SEQ ID 104>. This protein is predicted to be phoh-like protein (phoH). Analysis of this protein sequence reveals the following:

```
Possible site: 38

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2339(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB14476 GB:Z99117 phosphate starvation-induced protein
```

```
            [Bacillus subtilis]
 Identities = 191/305 (62%), Positives = 241/305 (78%), Gaps = 1/305 (0%)

Query: 27  LQHPDDMMSLFGSNERHLKLIEENLDVIIHARTERVQVLGDSEEAVETARLTIEALLVLV 86
           L++PD+ +SLFG+ +  LKL+E++L++ I  R E + V GD +E+ + A   + +LL L+
Sbjct: 12  LKNPDEALSLFGNQDSFLKLMEKDLNLNIITRGETIYVSGD-DESFQIADRLLGSLLALI 70

Query: 87  NRGMTVNTSDVVTALSMAQNGSIDKFVALYEEEIIKDSYGKPIRVKTLGQKIYVDSVKNH 146
            +G+ ++  DV+ A+ MA+   ++ F ++YEEEI K++ GK IRVKT+GQ+ YV ++K +
Sbjct: 71  RKGIEISERDVIYAIKMAKKNELEYFESMYEEEITKNAKGKSIRVKTMGQREYVAAMKRN 130

Query: 147 DVVFGIGPAGTGKTFLAVTLAVTALKRGQVKRIILTRPAVEAGESLGFLPGDLKEKVDPY 206
           D+VFGIGPAGTGKT+LAV  AV ALK G +K+IILTRPAVEAGESLGFLPGDLKEKVDPY
Sbjct: 131 DLVFGIGPAGTGKTYLAVVKAVHALKNGHIKKIILTRPAVEAGESLGFLPGDLKEKVDPY 190

Query: 207 LRPVYDALYQILGKEQTSRLMEREIIEIAPLAYMRGRTLDDAFVILDEAQNTTIMQMKMF 266
           LRP+YDAL+ +LG + T RLMER IIEIAPLAYMRGRTLDDA+VILDEAQNTT  QMKMF
Sbjct: 191 LRPLYDALHDVLGADHTERLMERGIIEIAPLAYMRGRTLDDAYVILDEAQNTTPAQMKMF 250

Query: 267 LTRLGFNSKMIVNGDVSQIDLPKNVKSGLIDAVEKLRNIKKIDFIHLSAKDVVRHPVVAE 326
           LTRLGF+SKMI+ GDVSQIDLPK VKSGL  A E L+ I  I  I L   DVVRHP+VA+
Sbjct: 251 LTRLGFSSKMIITGDVSQIDLPKGVKSGLAVAKEMLKGIDGISMIELDQTDVVRHPLVAK 310

Query: 327 IINAY 331
           II AY
Sbjct: 311 IIEAY 315
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 105> which encodes the amino acid sequence <SEQ ID 106>. Analysis of this protein sequence reveals the following:

```
Possible site: 42

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.85   Transmembrane   54 -  70 (  54 -  70)

----- Final Results -----
              bacterial membrane --- Certainty=0.1341(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 274/322 (85%), Positives = 298/322 (92%)

Query: 18  LQEYSIEITLQHPDDMMSLFGSNERHLKLIEENLDVIIHARTERVQVLGDSEEAVETARL 77
           LQEYSI+ITL HPDD+++LFGSNERHLKLIE +L VI+HARTERVQV+GD EEAVE ARL
Sbjct: 1   LQEYSIDITLTHPDDVLALFGSNERHLKLIEAHLGVIVHARTERVQVIGDDEEAVELARL 60

Query: 78  TIEALLVLVNRGMTVNTSDVVTALSMAQNGSIDKFVALYEEEIIKDSYGKPIRVKTLGQK 137
           TI+ALLVLV RGM VNTSDVVTALSMA++  ID+F+ALYEEEIIKD+YGK IRVKTLGQK
Sbjct: 61  TIKALLVLVGRGMVVNTSDVVTALSMAESHQIDQFMALYEEEIIKDNYGKAIRVKTLGQK 120

Query: 138 IYVDSVKNHDVVFGIGPAGTGKTFLAVTLAVTALKRGQVKRIILTRPAVEAGESLGFLPG 197
            YVDSVK HDVVFG+GPAGTGKTFLAVTLAVTALKRGQVKRIILTRPAVEAGESLGFLPG
Sbjct: 121 TYVDSVKRHDVVFGVGPAGTGKTFLAVTLAVTALKRGQVKRIILTRPAVEAGESLGFLPG 180

Query: 198 DLKEKVDPYLRPVYDALYQILGKEQTSRLMEREIIEIAPLAYMRGRTLDDAFVILDEAQN 257
           DLKEKVDPYLRPVYDALY ILGKEQT+RLMER++IEIAPLAYMRGRTLDDAFVILDEAQN
Sbjct: 181 DLKEKVDPYLRPVYDALYHILGKEQTTRLMERDVIEIAPLAYMRGRTLDDAFVILDEAQN 240

Query: 258 TTIMQMKMFLTRLGFNSKMIVNGDVSQIDLPKNVKSGLIDAVEKLRNIKKIDFIHLSAKD 317
           TTIMQMKMFLTRLGFNSKMIVNGD SQIDLP+NVKSGLIDA +KL+ IK+IDF++ SAKD
Sbjct: 241 TTIMQMKMFLTRLGFNSKMIVNGDTSQIDLPRNVKSGLIDATQKLQGIKQIDFVYFSAKD 300

Query: 318 VVRHPVVAEIINAYSDSESSHK 339
           VVRHPVVA+II AY  S    K
Sbjct: 301 VVRHPVVADIIKAYETSSEEMK 322
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 34

A DNA sequence (GBSx0032) was identified in *S.agalactiae* <SEQ ID 107> which encodes the amino acid sequence <SEQ ID 108>. Analysis of this protein sequence reveals the following:

```
Possible site: 30

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0275(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 35

A DNA sequence (GBSx0033) was identified in *S.agalactiae* <SEQ ID 109> which encodes the amino acid sequence <SEQ ID 110>. This protein is predicted to be MutT/nudix family protein. Analysis of this protein sequence reveals the following:

```
Possible site: 46

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2383(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF09597 GB:AE001864 MutT/nudix family protein [Deinococcus radiodurans]
 Identities = 49/136 (36%), Positives = 69/136 (50%), Gaps = 8/136 (5%)

Query: 5   YISYIRSKVGHETIFLTYSGGILTDGKGRVLLQLRADKNSWGIIGGCMELGESSVDTLKR 64
            Y+S +R+  GH  +       +L D  GRVLLQ R D   WGI+GG +E GE  +    R
Sbjct: 6   YLSELRAVWGHRALPAAGVSVLLQDETGRVLLQRRGDDGQWGILGGGLEPGEDFLIAAHR 65

Query: 65  EFFEETGLRVEPIRLLNVY------TNFQDSYPNGDKAQTVGFIYEVSCPKPVNIEGFHN 118
            E  EETGLR   +R L +         F   YPNGD+   VG   E + P     +   +
Sbjct: 66  ELLEETGLRCPNLRPLPLSEGLVSGPQFWHRYPNGDEVYLVGLRTEGTVPAAALTDACPD 125

Query: 119 E--ETLQLDYFSKEDV 132
            +  ETL+L +F+ +D+
Sbjct: 126 DGGETLELRWFALDDL 141
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 111> which encodes the amino acid sequence <SEQ ID 112>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4375(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 93/157 (59%), Positives = 123/157 (78%)

Query: 1   MKQDYISYIRSKVGHETIFLTYSGGILTDGKGRVLLQLRADKNSWGIIGGCMELGESSVD 60
           M QDYISYIRSKVGH+ I L ++GGILT+  G+VL+QLR DK +W I GG MELGESS++
Sbjct: 16  MPQDYISYIRSKVGHDKIILNFAGGILTNDDGKVLMQLRGDKKTWTIPGGTMELGESSLE 75

Query: 61  TLKREFFEETGLRVEPIRLLNVYTNFQDSYPNGDKAQTVGFIYEVSCPKPVNIEGFHNEE 120
           T KREF EETG+ VE +RLLNVYT+F++ YPNGD  QT+ FIYE++     + FHNEE
Sbjct: 76  TCKREFLEETGIEVEAVRLLNVYTHFEEVYPNGDAVQTIVFIYELTAVSDMAIDNFHNEE 135

Query: 121 TLQLDYFSKEDVKNITIVNEQHQLILDEYFSQTFQMG 157
           TL+L +FS E++  +  V+ +H+L+L+EYFS +F MG
Sbjct: 136 TLKLQFFSHEEIAELESVSAKHRLMLEEYFSDSFAMG 172
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 36

A DNA sequence (GBSx0034) was identified in *S.agalactiae* <SEQ ID 113> which encodes the amino acid sequence <SEQ ID 114>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3690(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 37

A DNA sequence (GBSx0035) was identified in *S.agalactiae* <SEQ ID 115> which encodes the amino acid sequence <SEQ ID 116>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAG05249 GB:AE004612 hypothetical protein [Pseudomonas aeruginosa]
 Identities = 70/254 (27%), Positives = 127/254 (49%), Gaps = 2/254 (0%)

Query: 2   KITLHGVAETLLITLYIRAKDAMAKHPILNDQKSLAIVEQIEYDFDKFDNSEASFYATLA 61
           +ITL G  +TLLITLY +A D+      IL+D+ +   V QI++DF +    + +  A
Sbjct: 5   RITLTGEKQTLLITLYAKALDSRLDDSILHDRFAEEAVRQIDFDFSRVALGKGNERALAM 64

Query: 62  RIRVMDREIKKFIRENPNSQILSIGCGLDTRFERVD-NGQIRWYNLDLPEVMEIRKLFFE 120
           R    D+  ++F+  +P  Q+L++GCGLD+R  RVD    ++ W++LD PEVM++R+  +
Sbjct: 65  RSHYFDQACREFLGRHPEGQVLNLGCGLDSRIYRVDPPAELPWFDLDYPEVMDLRERLYP 124

Query: 121 EHERVTNIAKSALDETWTREVNPQNAPFLIVSEGVLMFLKEDDVETFLHILTNSFSQFMA 180
                   + ++D+    +  P+  P L+++EG++ +L+E  V   +  L +
Sbjct: 125 PRAGAYRALRHSVDDDGWLQGVPRERPALVLAEGLMPYLRESQVRRLVERLVDHLGSGEL 184

Query: 181 QFDLCHKEMINKGKQHDTVKYMDTEFQFGITDGHEIVDLDPKLKQINLINFTDEMSKFEL 240
            FD   +  I   + +  ++    +  + I D  E+    P L+ I  +  D     +L
Sbjct: 185 LFDGYGRLGIMLLRLYPPLRETGAQVHWSIDDPRELERWHPALRFIEEVTDYDPQDVAKL 244

Query: 241 -GTLRSLLPTIRKF 253
             + R +LP    F
Sbjct: 245 PQSSRLMLPIYNGF 258
```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8477> and protein <SEQ ID 8478> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 9
McG: Discrim Score:      0.37
GvH: Signal Score (-7.5): -0.97
     Possible site: 25
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:   4.35 threshold:  0.0
   PERIPHERAL  Likelihood =  4.35     143
 modified ALOM score:  -1.37

*** Reasoning Step: 3

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
27.6/51.6% over 253aa
                                                          Pseudomonas aeruginosa
  GP|9947849| hypothetical protein Insert characterized

ORF02096(304 - 1059 of 1404)
GP|9947849|gb|AAG05249.1|AE004612_3|AE004612(5   -   258   of   275)   hypothetical   protein
{Pseudomonas aeruginosa}
%Match = 11.6
%Identity = 27.6  %Similarity = 51.6
Matches = 70  Mismatches = 121  Conservative Sub.s = 61

255       285       315       345       375       405       435       465
E*YT*RNPVLEIQISK*NSIKESR*MKITLHGVAETLLITLYIRAKDAMAKHPILNDQKSLAIVEQIEYDFDKFDNSEAS
                         :||| |  :|||||| :| |:     ||:|: :   | ||::|| :    : :
                       MPGHRITLTGEKQTLLITLYAKALDSRLDDSILHDRFAEEAVRQIDFDFSRVALGKGN
                               10        20        30        40        50

495       525       555       585       612       642       672       702
FYATLARXRVMDREIKKFIRENPNSQILSIGCGLDTRFERVDN-GQIRWYNLDLPEVMEIRKLFFEEHERVTNIAKSALD
  |   |    |:  ::|:  :|  |:|::|||||:|  |||   :: |::|| ||||::|: ::          : ::|
ERALAMRSHYFDQACREFLGRHPEGQVLNLGCGLDSRIYRVDPPAELPWFDLDYPEVMDLRERLYPPRAGAYRALRHSVD
```

```
          70        80        90        100       110       120       130

732       762       792       822       852       882       912       942
ETWTREVNPQNAPFLIVSEGVLMFLKEDDVETFLHILTNSFSQFMAQFDLCHKEMINKGKQHDTVKYMDTEFQFGITDGH
:    :  |:  | |:::||:: :|:|  |  ::  | : :      ||   :  |   : :  ::    : :: | |
DDGWLQGVPRERPALVLAEGLMPYLRESQVRRLVERLVDHLGSGELLFDGYGRLGIMLLRLYPPLRETGAQVHWSIDDPR
          150       160       170       180       190       200       210

972       1002      1029      1059      1089      1119      1149      1179
EIVDLDPKLKQINLINFTDEMSKFELG-TLRSLLPTIRKFNNCLGVYEYKASEKK*QKSIYIKRHSKCKFVIIVIAFVAL
|:    | |: |  :   |     :|  : | :||    |     :
ELERWHPALRFIEEVTDYDPQDVAKLPQSSRLMLPIYNGFAFLRRMGRLIRYRWPRV
          230       240       250       260       270
```

SEQ ID 8478 (GBS176) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 36 (lane 5 & 6; MW 30kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 41 (lane 7; MW 55.4kDa).

The GBS176-GST fusion product was purified (Figure 117A; see also Figure 202, lane 5) and used to immunise mice (lane 1+2 product; 13.5μg/mouse). The resulting antiserum was used for Western blot (Figure 117B), FACS (Figure 117C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 38

A DNA sequence (GBSx0036) was identified in *S.agalactiae* <SEQ ID 117> which encodes the amino acid sequence <SEQ ID 118>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3712(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10019> which encodes amino acid sequence <SEQ ID 10020> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC38046 GB:AF000954 No definition line found [Streptococcus mutans]
 Identities = 140/164 (85%), Positives = 157/164 (95%)

Query: 1   MYVEMIDETGQVSEDIKKQTLDLLEFAAQKTGKENKEMAVTFVTNERSHELNLEYRDTDR 60
           MY+EMIDET QVSE IK QTLD+LEFAAQKTGKE+KEMAVTFVTNERSHELNL+YRDT+R
Sbjct: 1   MYIEMIDETNQVSEGIKNQTLDILEFAAQKTGKEDKEMAVTFVTNERSHELNLKYRDTNR 60

Query: 61  PTDVISLEYKPEVDISFDEEDLAENPELAEMLEDFDSYIGELFISIDKAKEQAEEYGHSY 120
           PTDVISLEYKPE  +SFDEEDLA++P+LAE+L +FD+YIGELFIS+DKA+EQA+EYGHS+
Sbjct: 61  PTDVISLEYKPESSLSFDEEDLADDPDLAEVLTEFDAYIGELFISVDKAREQAQEYGHSF 120

Query: 121 EREMGFLAVHGFLHINGYDHYTPEEEKEMFSLQEEILTAYGLKR 164
           EREMGFLAVHGFLHINGYDHYTP+EEKEMFSLQEEIL AYGLKR
Sbjct: 121 EREMGFLAVHGFLHINGYDHYTPQEEKEMFSLQEEILDAYGLKR 164
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 119> which encodes the amino acid sequence <SEQ ID 120>. Analysis of this protein sequence reveals the following:

```
Possible site: 49

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1145(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 138/165 (83%), Positives = 153/165 (92%)

Query: 1   MYVEMIDETGQVSEDIKKQTLDLLEFAAQKTGKENKEMAVTFVTNERSHELNLEYRDTDR 60
           MY+EMIDETGQVS++I +QTLDLL FAAQKTGKE KEM+VTFVTNERSHELNLEYRDTDR
Sbjct: 18  MYIEMIDETGQVSQEIMEQTLDLLNFAAQKTGKEEKEMSVTFVTNERSHELNLEYRDTDR 77

Query: 61  PTDVISLEYKPEVDISFDEEDLAENPELAEMLEDFDSYIGELFISIDKAKEQAEEYGHSY 120
           PTDVISLEYKPE  I F +EDLA +P LAEM+ +FD+YIGELFISIDKA+EQ++EYGHS+
Sbjct: 78  PTDVISLEYKPETPILFSQEDLAADPSLAEMMAEFDAYIGELFISIDKAREQSQEYGHSF 137

Query: 121 EREMGFLAVHGFLHINGYDHYTPEEEKEMFSLQEEILTAYGLKRQ 165
           EREMGFLAVHGFLHINGYDHYT EEEKEMF+LQEEILTAYGL RQ
Sbjct: 138 EREMGFLAVHGFLHINGYDHYTLEEEKEMFTLQEEILTAYGLTRQ 182
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 39

A DNA sequence (GBSx0038) was identified in *S.agalactiae* <SEQ ID 121> which encodes the amino acid sequence <SEQ ID 122>. This protein is predicted to be phosphoglycerate dehydrogenase (serA) (serA). Analysis of this protein sequence reveals the following:

```
Possible site: 59

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2817(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB99020 GB:U67544 phosphoglycerate dehydrogenase (serA)
           [Methanococcus jannaschii]
 Identities = 82/232 (35%), Positives = 132/232 (56%), Gaps = 14/232 (6%)

Query: 3   ENPDAYIIRSQNLHNQDF---PSNLKAIARAGAGTNNIPIEEASAQGIVVFNTPGANANA 59
           ++ D  ++RS     +D       LK I RAG G +NI +E A+ +GI+V N P A++ +
Sbjct: 40  KDADVLVVRSGTKVTRDVIEKAEKLKVIGRAGVGVDNIDVEAATEKGIIVVNAPDASSIS 99

Query: 60  VKEAVIAALLLSARDYLGANRWVNTLTGTDIPKQIEAGKKAFAGNEIAGKKLGVIGLGAI 119
           V E  +  +L +AR         N    T   K+ E  +K F G E+ GK LGVIGLG I
Sbjct: 100 VAELTMGLMLAAAR---------NIPQATASLKRGEWDRKRFKGIELYGKTLGVIGLGRI 150

Query: 120 GARIANDARRLGMTVLGYDPYVSIETAWNISSHVQRVKEIKDIFETCDYITIHVPLTNET 179
           G ++   A+  GM ++GYDPY+  E A ++   V+ V +I ++ +  D+IT+HVPLT +T
Sbjct: 151 GQQVVKRAKAFGMNIIGYDPYIPKEVAESMG--VELVDDINELCKRADFITLHVPLTPKT 208
```

```
Query: 180 KHTFDAKAFSIMKKGTTIINFARAELVNNQELFEAIETGVVKRYITDFGDKE 231
           +H    +  ++MKK   I+N AR  L++ + L+EA++ G ++    D  ++E
Sbjct: 209 RHIIGREQIALMKKNAIIVNCARGGLIDEKALYEALKEGKIRAAALDVFEEE 260
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 123> which encodes the amino acid sequence <SEQ ID 124>. Analysis of this protein sequence reveals the following:

```
Possible site: 52

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2384(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 52/198 (26%), Positives = 93/198 (46%), Gaps = 14/198 (7%)

Query: 24  LKAIARAGAGTNNIPIEEASAQGIVVFNTPGANANAVKEAVIAALLLSARDYLGANRWVN 83
           +K IA+  A  +   ++ A+   I++ N P  +  ++ E  +  +L   R
Sbjct: 70  IKQIAQHSASVDMYNLDLATENDIIITNVPSYSPESIAEFTVTIVLNLIRHV-------- 121

Query: 84  TLTGTDIPKQIEAGKKAFAGNEIAGKKLGVIGLGAIGARIANDARRLGMTVLGYDPYVSI 143
            L   ++ KQ         G  +    + +IG G IG   A   +  G  V+GYD Y S
Sbjct: 122 ELIRENVKKQNFTWGLPIRGRVLGDMTVAIIGTGRIGLATAKIFKGFGCKVVGYDIYQS- 180

Query: 144 ETAWNISSHVQRVKE-IKDIFETCDYITIHVPLTNETKHTFDAKAFSIMKKGTTIINFAR 202
           + A  +  + + V+E IKD     D +++H+P T E  H F++  F   KKG  ++N AR
Sbjct: 181 DAAKAVLDYKESVEEAIKD----ADLVSLHMPPTAENTHLFNSDLFKSFKKGAILMNMAR 236

Query: 203 AELVNNQELFEAIETGVV 220
             ++  Q+L +A++ G++
Sbjct: 237 GAVIETQDLLDALDAGLL 254
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 40

A DNA sequence (GBSx0039) was identified in *S.agalactiae* <SEQ ID 125> which encodes the amino acid sequence <SEQ ID 126>. This protein is predicted to be alpha-glycerophosphate oxidase. Analysis of this protein sequence reveals the following:

```
Possible site: 50

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2067(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC34740 GB:U94770 alpha-glycerophosphate oxidase [Streptococcus pneumoniae]
 Identities = 24/49 (48%), Positives = 37/49 (74%)

Query: 1   MLFMRDNLDSLIQPVIDEMAKHYQWSDQDKTFYEEELHETLKDNDLAAL 49
           MLFMRD+LDS+++PV+DEM + Y W++++K  Y  ++   L +NDLA L
Sbjct: 558 MLFMRDSLDSIVEPVLDEMGRFYDWTEEEKATYRADVEAALANNDLAEL 606
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 127> which encodes the amino acid sequence <SEQ ID 128>. Analysis of this protein sequence reveals the following:

```
     Possible site: 40
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.81   Transmembrane   20 -  36 (  20 -  36)

----- Final Results -----
               bacterial membrane --- Certainty=0.1723(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC34740 GB:U94770 alpha-glycerophosphate oxidase [Streptococcus pneumoniae]
 Identities = 462/607 (76%), Positives = 539/607 (88%)

Query: 1   MEFSRETRRLALQKMQERDLDLLIIGGGITGAGVALQAAASGLDTGLIEMQDFAQGTSSR 60
           MEFS++TR L+++KMQER LDLLIIGGGITGAGVALQAAASGL+TGLIEMQDFA+GTSSR
Sbjct: 1   MEFSKKTRELSIKKMQERTLDLLIIGGGITGAGVALQAAASGLETGLIEMQDFAEGTSSR 60

Query: 61  STKLVHGGLRYLKQFDVEVVSDTVSERAVVQQIAPHIPKPDPMLLPVYDEPGSTFSMFRL 120
           STKLVHGGLRYLKQFDVEVVSDTVSERAVVQQIAPHIPKPDPMLLPVYDE G+TFS+FRL
Sbjct: 61  STKLVHGGLRYLKQFDVEVVSDTVSERAVVQQIAPHIPKPDPMLLPVYDEDGATFSLFRL 120

Query: 121 KVAMDLYDLLAGVSNTPAANKVLTKEEVLKREPDLKQEGLLGGGVYLDFRNNDARLVIEN 180
           KVAMDLYDLLAGVSNTP ANKVL+K++VL+R+P+LK+EGL+GGGVYLDFRNNDARLVIEN
Sbjct: 121 KVAMDLYDLLAGVSNTPTANKVLSKDQVLERQPNLKKEGLVGGGVYLDFRNNDARLVIEN 180

Query: 181 IKRANRDGALIASHVKAEDFLLDDNGKIIGVKARDLLSDQEIIIKAKLVINTTGPWSDEI 240
           IKRAN+DGALIA+HVKAE FL D++GKI GV ARDLL+DQ   IKA+LVINTTGPWSD++
Sbjct: 181 IKRANQDGALIANHVKAEGFLFDESGKITGVVARDLLTDQVFEIKARLVINTTGPWSDKV 240

Query: 241 RQFSHKGQPIHQMRPTKGVHLVVDRQKLPVSQPVYVDTGLNDGRMVFVLPREEKTYFGTT 300
           R  S+KG    QMRPTKGVHLVVD  K+ VSQPVY DTGL DGRMVFVLPRE KTYFGTT
Sbjct: 241 RNLSNKGTQFSQMRPTKGVHLVVDSSKIKVSQPVYFDTGLGDGRMVFVLPRENKTYFGTT 300

Query: 301 DTDYTGDLEHPQVTQEDVDYLLGVVNNRFPNANVTIDDIESSWAGLRPLLSGNSASDYNG 360
           DTDYTGDLEHP+VTQEDVDYLLG+VNNRFP +N+TIDDIESSWAGLRPL++GNSASDYNG
Sbjct: 301 DTDYTGDLEHPKVTQEDVDYLLGIVNNRFPESNITIDDIESSWAGLRPLIAGNSASDYNG 360

Query: 361 GNSGKVSDDSFDHLVDTVKAYINHEDSREAVEKAIKQVETSTSEKELDPSAVSRGSSFER 420
           GN+G +SD+SFD+L+ TV++Y++ E +RE VE A+ ++E+STSEK LDPSAVSRGSS +R
Sbjct: 361 GNNGTISDESFDNLIATVESYLSKEKTREDVESAVSKLESSTSEKHLDPSAVSRGSSLDR 420

Query: 421 DENGLFTLAGGKITDYRKMAEGALTGIIQILKEEFGKSFKLINSKTYPVSGGEINPANVD 480
           D+NGL TLAGGKITDYRKMAEGA+  ++ ILK EF +SFKLINSKTYPVSGGE+NPANVD
Sbjct: 421 DDNGLLTLAGGKITDYRKMAEGAMERVVDILKAEFDRSFKLINSKTYPVSGGELNPANVD 480

Query: 481 SEIEAYAQLGTLSGLSMDDARYLANLYGSNAPKVFALTRQLTAAEGLSLAETLSLHYAMD 540
           SEIEA+AQLG   GL   +A YLANLYGSNAPKVFAL   L  A GLSLA+TLSLHYAM
Sbjct: 481 SEIEAFAQLGVSRGLDSKEAHYLANLYGSNAPKVFALAHSLEQAPGLSLADTLSLHYAMR 540

Query: 541 YEMALKPTDYFLRRTNHLLFMRDSLDALIDPVINEMAKHFEWSDQERVAQEDDLRRVIAD 600
            E+AL P D+ LRRTNH+LFMRDSLD++++PV++EM + ++W+++E+     D+   +A+
Sbjct: 541 NELALSPVDFLLRRTNHMLFMRDSLDSIVEPVLDEMGRFYDWTEEEKATYRADVEAALAN 600

Query: 601 NDLSALK 607
           NDL+ LK
Sbjct: 601 NDLAELK 607
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 29/49 (59%), Positives = 41/49 (83%)

Query: 1   MLFMRDNLDSLIQPVIDEMAKHYQWSDQDKTFYEEELHETLKDNDLAAL 49
           +LFMRD+LD+LI PVI+EMAKH++WSDQ++   E++L   + DNDL+AL
```

```
Sbjct: 558 LLFMRDSLDALIDPVINEMAKHFEWSDQERVAQEDDLRRVIADNDLSAL 606
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 41

A DNA sequence (GBSx0040) was identified in *S.agalactiae* <SEQ ID 129> which encodes the amino acid sequence <SEQ ID 130>. Analysis of this protein sequence reveals the following:

```
Possible site: 40

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1011(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB06309 GB:AP001516 unknown conserved protein [Bacillus halodurans]
 Identities = 70/160 (43%), Positives = 106/160 (65%), Gaps = 3/160 (1%)

Query: 5   TRPTTDKVKGAIFNMIGPFFEGGRVLDLFSGSGSLAIEAISRGMDQAVLVEKDRRAQVVI 64
           TRPTTDKVK AIFNMIGPFF+GG  LDL+ GSG L IEA+SRG+++ + V++ +RA   I
Sbjct: 21  TRPTTDKVKEAIFNMIGPFFDGGIGLDLYGGSGGLGIEALSRGVERMIFVDQQKRAIETI 80

Query: 65  QENIAMTKSPEQFQLLKMEANRALEQLTGQ---FDLVLLDPPYAKEEIVKQIQIMDSKGL 121
           ++N++       + ++ + +A RAL+ LT +    F  V LDPPYAK+ I    + I+ + GL
Sbjct: 81  KQNLSHCGLEGRAEVYRNDAKRALQVLTKRGIVFAYVFLDPPYAKQTIKNDLAILANHGL 140

Query: 122 LGDDIMIACETDKSVDLPEEIASFGIWKQKIYGISKVTVY 161
           L +  ++ CE D+   LP++I      K++ YG + +T+Y
Sbjct: 141 LEEGGVVVCEHDRDTMLPDQIEYAVKHKEETYGDTMITIY 180
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 131> which encodes the amino acid sequence <SEQ ID 132>. Analysis of this protein sequence reveals the following:

```
Possible site: 58

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3814(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 111/160 (69%), Positives = 136/160 (84%)

Query: 3   RTTRPTTDKVKGAIFNMIGPFFEGGRVLDLFSGSGSLAIEAISRGMDQAVLVEKDRRAQV 62
           + TRPT+DKV+GAIFNMIGP+F GGRVLDLF+GSG LAIEA+SRGM  AVLVEK+R+AQ
Sbjct: 19  KITRPTSDKVRGAIFNMIGPYFNGGRVLDLFAGSGGLAIEAVSRGMSAAVLVEKNRKAQA 78

Query: 63  VIQENIAMTKSPEQFQLLKMEANRALEQLTGQFDLVLLDPPYAKEEIVKQIQIMDSKGLL 122
           +IQ+NI MTK+  +F LLKMEA RA++ LTG+FDLV LDPPYAKE IV  I+ + +K LL
Sbjct: 79  IIQDNIIMTKAENRFTLLKMEAERAIDCLTGRFDLVFLDPPYAKETIVATIEALAAKNLL 138

Query: 123 GDDIMIACETDKSVDLPEEIASFGIWKQKIYGISKVTVYV 162
            + +M+ CETDK+V LP+EIA+ GIWK+KIYGISKVTVYV
Sbjct: 139 SEQVMVVCETDKTVLLPKEIATLGIWKEKIYGISKVTVYV 178
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 42

A DNA sequence (GBSx0041) was identified in *S.agalactiae* <SEQ ID 133> which encodes the amino acid sequence <SEQ ID 134>. This protein is predicted to be lipopolysaccharide core biosynthesis protein kdtB (kdtB). Analysis of this protein sequence reveals the following:

```
Possible site: 17

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1937(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB13272 GB:AP001119 lipopolysaccharide core biosynthesis
           protein kdtB [Buchnera sp. APS]
 Identities = 56/149 (37%), Positives = 94/149 (62%)

Query: 1   MTKKALFTGSFDPVTNGHLDIIERASYLFDHVYIGLFYNLEKQGYFSIECRKKMLEEAIR 60
           M K A++ G+FDP+T GHLDII RA+ +FD + I +  N  K+  F+++ R ++  +
Sbjct: 1   MNKTAIYPGTFDPITYGHLDIITRATKIFDSITIAISNNFTKKPIFNLKERIELTRKVTL 60

Query: 61  QFKNVSVLVAQDRLAVDLAREVGAKYFVRGLRNSQDFDYEANLEFFNKQLADDIETVYLS 120
             KNV  ++  + L  +LA++  A   +RG+R   DFDYE  L   NKQ+  D+++++L
Sbjct: 61  HLKNVKKILGFNDLLANLAKKEKANILIRGVRTIFDFDYEIKLAAINKQIYPDLDSIFLL 120

Query: 121 TSPSLSPISSSRIRELIHFKASVKPFVPK 149
           +S  +S ISSS ++E+  +K  +KP++PK
Sbjct: 121 SSKEVSFISSSFVKEIAKYKGDIKPYLPK 149
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 135> which encodes the amino acid sequence <SEQ ID 136>. Analysis of this protein sequence reveals the following:

```
Possible site: 61

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1862(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 88/161 (54%), Positives = 124/161 (76%)

Query: 1   MTKKALFTGSFDPVTNGHLDIIERASYLFDHVYIGLFYNLEKQGYFSIECRKKMLEEAIR 60
           +TK  L+TGSFDPVTNGHLDI++RAS LFD +Y+G+F N  K+ YF +E RK ML +A+
Sbjct: 2   LTKIGLYTGSFDPVTNGHLDIVKRASGLFDQIYVGIFDNPTKKSYFKLEVRKAMLTQALA 61

Query: 61  QFKNVSVLVAQDRLAVDLAREVGAKYFVRGLRNSQDFDYEANLEFFNKQLADDIETVYLS 120
            F NV V+ + +RLA+D+A+E+    + +RGLRN+ DF+YE NLE+FN  LA +IETVYL
Sbjct: 62  DFTNVIVVTSHERLAIDVAKELRVTHLIRGLRNATDFEYEENLEYFNHLLAPNIETVYLI 121

Query: 121 TSPSLSPISSSRIRELIHFKASVKPFVPKSVVREVEKMSEE 161
           +      +SSSR+RELIHF++S++  VP+SV+ +VEKM+E+
Sbjct: 122 SRNKWQALSSSRVRELIHFQSSLEGLVPQSVIAQVEKMNEK 162
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 43

A DNA sequence (GBSx0042) was identified in *S.agalactiae* <SEQ ID 137> which encodes the amino acid sequence <SEQ ID 138>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1126(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 44

A DNA sequence (GBSx0043) was identified in *S.agalactiae* <SEQ ID 139> which encodes the amino acid sequence <SEQ ID 140>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.04   Transmembrane   20 -  36 (  12 -  43)

----- Final Results -----
               bacterial membrane --- Certainty=0.5416(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB13378 GB:Z99111 ylbL [Bacillus subtilis]
 Identities = 124/344 (36%), Positives = 199/344 (57%), Gaps = 21/344 (6%)

Query: 20  WIIGFAFLLLVLASLVVRLPYYLEMPGGAYDIRSVLKVNKKADKAKGSYNFVAVSVSQAT 79
           W++    L+ VL+   ++LPYY+  PG A ++ S++KV     + KGS + + V V  A
Sbjct: 9   WMLVILILIAVLS--FIKLPYYITKPGEATELASLIKVEGGYPE-KGSLSLMTVKVGPAN 65

Query: 80  PAQVLYAWLTPFTEL----SSKEETTGGFSNDDYLRINQFYMETSQNESIYQALKLANKQ 135
           P   ++A + P+ E+    S KEE   G S+ +Y++     M++SQ  ++  A + A K+
Sbjct: 66  PFTYVWAKMHPYYEIVPDESIKEE---GESDKEYMKRQLQMMKSSQENAVIAAYQKAGKK 122

Query: 136 VSLTYKGVYVLNLAKNSTFKDRLHLADTVTGVNGKSFKNSSQLIKYVAALHLGDKVKVQY 195
           VS ++ G+Y  ++ +N   K ++ + D +   +GK+++++ +LI Y+++   GDKV ++
Sbjct: 123 VSYSFNGIYASSVVENMPAKGKIEVGDKIISADGKNYQSAEKLIDYISSKKAGDKVTLKI 182

Query: 196 TSQGKKKESVGKVIKLSNGKNGIGIGLTDHTE--VLSDVPVDFNTEGVGGPSAGLMFTLA 253
             + K+K     + +  +  GIG++ +T+  V  +  +DF  E +GGPSAGLM +L
Sbjct: 183 EREEKEKRVTLTLKQFPDEPDRAGIGVSLYTDRNVKVEPDIDFEIENIGGPSAGLMMSLE 242

Query: 254 IYDQLVKEDLRKGRKIAGTGTIEQNGHVGDIGGAGLKVVSAAKKGMDIFFVPNNPIDKNA 313
           IY+QL K D  KG  IAGTGTI+ +G VG IGG   KVV+A K G DIFF PN     N
Sbjct: 243 IYNQLTKPDETKGYDIAGTGTIDVDGKVGPIGGIDQKVVAADKAGKDIFFAPNQNGASN- 301
```

```
Query: 314 KKGKTKVQTNYQEAKAAAKRLGTKMKIVPVQNVQQAIDYLKKTK 357
                  ++Y+ A   AK + + MKIVPV  +Q AIDYL K K
Sbjct: 302 --------SDYKNAVKTAKDIDSNMKIVPVDTMQDAIDYLNKLK 337
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 141> which encodes the amino acid sequence <SEQ ID 142>. Analysis of this protein sequence reveals the following:

```
     Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.24   Transmembrane   10 -  26 (   6 -  34)

----- Final Results -----
               bacterial membrane --- Certainty=0.5097(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB13378 GB:Z99111 ylbL [Bacillus subtilis]
 Identities = 132/348 (37%), Positives = 198/348 (55%), Gaps = 16/348 (4%)

Query: 1   MKRLKKIKWWLVGLLALISLLLALFFPLPYYIEMPGGAYDIRTVLQVNGKEDKRKGAYQF 60
           M R K   W LV +L LI++L   F  LPYYI  PG A ++ ++++V G   + KG+
Sbjct: 1   MLRKKHFSWMLV-ILILIAVLS--FIKLPYYITKPGEATELASLIKVEGGYPE-KGSLSL 56

Query: 61  VAVGISRASLAQLLYAWLTPFTEISTAEDTTG-GYSDADFLRINQFYMETSQNAAIYQAL 119
           + V +  A+    ++A + P+ EI   E     G SD ++++     M++SQ  A+  A
Sbjct: 57  MTVKVGPANPFTYVWAKMHPYYEIVPDESIKEEGESDKEYMKRQLQMMKSSQENAVIAAY 116

Query: 120 SLAGKPVTLDYKGVYVLDVNNESTFKGTLHLADTVTGVNGKQFTSSAELIDYVSHLKLGD 179
             AGK V+  + G+Y   V      KG + + D +   +GK + S+ +LIDY+S  K GD
Sbjct: 117 QKAGKKVSYSFNGIYASSVVENMPAKGKIEVGDKIISADGKNYQSAEKLIDYISSKKAGD 176

Query: 180 EVTVQFTSDNKPKKGVGRIIKLKN--GKNGIGIALTDHTSVNSEDTVIFSTKGVGGPSAG 237
           +VT++   + K K+     + +   + GIG++L     +V  E  + F  +GGPSAG
Sbjct: 177 KVTLKIEREEKEKRVTLTLKQFPDEPDRAGIGVSLYTDRNVKVEPDIDFEIENIGGPSAG 236

Query: 238 LMFTLDIYDQITKEDLRKGRTIAGTGTIGKDGEVGDIGGAGLKVVAAAEAGADIFFVPNN 297
           LM +L+IY+Q+TK D  KG  IAGTGTI  DG+VG IGG   KVVAA +AG DIFF PN
Sbjct: 237 LMMSLEIYNQLTKPDETKGYDIAGTGTIDVDGKVGPIGGIDQKVVAADKAGKDIFFAPNQ 296

Query: 298 PVDKEIKKVNPNAISNYEEAKRAAKRLKTKMKIVPVTTVQEALVYLRK 345
                N  + S+Y+ A + AK + + MKIVPV T+Q+A+ YL K
Sbjct: 297 ---------NGASNSDYKNAVKTAKDIDSNMKIVPVDTMQDAIDYLNK 335
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 229/339 (67%), Positives = 276/339 (80%)

Query: 17  LKWWIIGFAFLLLVLASLVVRLPYYLEMPGGAYDIRSVLKVNKKADKAKGSYNFVAVSVS 76
           +KWW++G   L+ +L +L   LPYY+EMPGGAYDIR+VL+VN K DK KG+Y FVAV +S
Sbjct: 7   IKWWLVGLLALISLLLALFFPLPYYIEMPGGAYDIRTVLQVNGKEDKRKGAYQFVAVGIS 66

Query: 77  QATPAQVLYAWLTPFTELSSKEETTGGFSNDDYLRINQFYMETSQNESIYQALKLANKQV 136
           +A+ AQ+LYAWLTPFTE+S+ E+TTGG+S+ D+LRINQFYMETSQN +IYQAL LA K V
Sbjct: 67  RASLAQLLYAWLTPFTEISTAEDTTGGYSDADFLRINQFYMETSQNAAIYQALSLAGKPV 126

Query: 137 SLTYKGVYVLNLAKNSTFKDRLHLADTVTGVNGKSFKNSSQLIKYVAALHLGDKVKVQYT 196
           +L YKGVYVL++   STFK  LHLADTVTGVNGK F +S++LI YV+ L LGD+V VQ+T
Sbjct: 127 TLDYKGVYVLDVNNESTFKGTLHLADTVTGVNGKQFTSSAELIDYVSHLKLGDEVTVQFT 186

Query: 197 SQGKKKESVGKVIKLSNGKNGIGIGLTDHTEVLSDVPVDFNTEGVGGPSAGLMFTLAIYD 256
           S  K K+ VG++IKL NGKNGIGI LTDHT V S+  V F+T+GVGGPSAGLMFTL IYD
Sbjct: 187 SDNKPKKGVGRIIKLKNGKNGIGIALTDHTSVNSEDTVIFSTKGVGGPSAGLMFTLDIYD 246

Query: 257 QLVKEDLRKGRKIAGTGTIEQNGHVGDIGGAGLKVVSAAKKGMDIFFVPNNPIDKNAKKG 316
           Q+ KEDLRKGR IAGTGTI ++G VGDIGGAGLKVV+AA+ G DIFFVPNNP+DK  KK
```

```
Sbjct: 247 QITKEDLRKGRTIAGTGTIGKDGEVGDIGGAGLKVVAAAEAGADIFFVPNNPVDKEIKKV 306

Query: 317 KTKVQTNYQEAKAAAKRLGTKMKIVPVQNVQQAIDYLKK 355
               +NY+EAK AAKRL TKMKIVPV  VQ+A+ YL+K
Sbjct: 307 NPNAISNYEEAKRAAKRLKTKMKIVPVTTVQEALVYLRK 345
```

A related GBS gene <SEQ ID 8479> and protein <SEQ ID 8480> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:      8.26
GvH: Signal Score (-7.5): -4.04
     Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 1 value: -11.04 threshold:  0.0
   INTEGRAL    Likelihood =-11.04   Transmembrane   20 -  36 (  12 -  43)
   PERIPHERAL  Likelihood =  4.51      70
 modified ALOM score:   2.71

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5416(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
GP|5531383| putative secreted protein {Streptomyces coelicolor A3(2)} Insert characterized
  PIR|T36157|T36157  probable  secreted  protein  -  Streptomyces  coelicolor  Insert
characterized

ORF01344(361 - 1362 of 1671)
GP|5531383|emb|CAB51015.1||AL096852(13  -  247  of  259)  putative  secreted  protein
{Streptomyces coelicolor A3(2)} PIR|T36157|T36157 probable secreted protein - Streptomyces
coelicolor
%Match = 7.1
%Identity = 38.4  %Similarity = 57.6
Matches = 58  Mismatches = 61  Conservative Sub.s = 29

312       342       372       402       432       462       492
EKWRK*VKNRDPKRKHKSLLGLLKWWIIGFAFLLLVLASLVVRLPYYLEMPGGAYDIRSVLKVNKKADKAKGSYNFV~~~
                |       : |:  : |:  :|   ||: :  ||   |:
              MLSRLTRPQFLAVCGLPVVALLATALFAPLPFSVAQPGLTADV----------------------~~~
                       10        20        30        40

924       954       984                                                  1002
~KKKESVGKVIKLSNGKNGIGIGLTDHTEVLS-------------------------~~~~~~~~~~~~~~~~~~~~~~~~~DVPV
                      :|     ||::                                                  || |
~---------------------LGKNRGAEVITISGAPTHATSGQLRMTTIEA~~~~KESQDSATTAALRYLRMDKGDVDV
                           50        60        70              130       140

1032      1062      1092      1122      1152      1182      1212      1242
DFNTEGVGGPSAGLMFTLAIYDQLVKEDLRKGRKIAGTGTIEQNGHVGDIGGAGLKVVSAAKKGMDIFFVPNNPIDKNAK
 :  | ||||||||:|:| | |:|   ||  |: :||||||   | || :||  ||  :| : |  :|:||
KLRLEDVGGPSAGLLFSLGIVDKLGAGDLTGGKVVAGTGTITDGGKVGAVGGVPLKTQAARRDGATVFLVPK--------
             160       170       180       190       200       210

1272      1302      1332      1362      1392      1422      1452      1482
KGKTKVQTNYQEAKAAAKRLGTKMKIVPVQNVQQAIDYLKKTK*TQRVRASARLFCFATFDYQSAKMIV*QSL*EYYI*M
          |   |   |   ::::||  :: |:| ||     :    :
----------AECSDAQAELPKGLRLIPVTTLEGAVDSLKALESGKGDVPAC
          220       230       240       250
```

SEQ ID 8480 (GBS39) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 12 (lane 9; MW 65.2kDa) and Figure 15 (lane 3; MW 40kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 45

A DNA sequence (GBSx0044) was identified in *S.agalactiae* <SEQ ID 143> which encodes the amino acid sequence <SEQ ID 144>. This protein is predicted to be UDP-sugar hydrolase. Analysis of this protein sequence reveals the following:

```
Possible site: 17

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3908(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB15227 GB:Z99120 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 114/280 (40%), Positives = 173/280 (61%), Gaps = 9/280 (3%)

Query: 1   MTELIRILHLNDLHSHFENFPKVKRFFH----DNQAQPIETISLDLGDNIDKSHPLTEAS 56
            M E +R+ H NDLHSHFEN+PK+  +       ++Q+   ET+  D+GD++D+   +TEA+
Sbjct: 1   MKEKLRLYHTNDLHSHFENWPKIVDYIEQKRKEHQSDGEETLVFDIGDHLDRFQFVTEAT 60

Query: 57  SGKANVQLMNELGIELATIGNNEGVGLSKKDLDQVYKDSDFTVIVGNLKD-NIIEPSWAK 115
             GKANV L+N L I+ A IGNNEG+ L  ++L  +Y  ++F VIV NL D N   PSWA
Sbjct: 61  FGKANVDLLNRLHIDGAAIGNNEGITLPHEELAALYDHAEFPVIVSNLFDKNGNRPSWAV 120

Query: 116 PYIIYETQQGTKLAFLAYTFPYYKTYEPNGWTIEDPIDCLKCHLQINEIK-EANCRILMS 174
            PY I   + G  +AFL  T PYY  Y+  GWT+ D ++ +K    I E+K +A+  +L+S
Sbjct: 121 PYHIKSLKNGMSIAFLGVTVPYYPVYDKLGWTVTDALESIK--ETILEVKGQADIIVLLS 178

Query: 175 HLGIRFDTRIAQEFSEIDLIIGAHTHHLFEEGELINGTYLAAAGKYGRFVGSIDITFDNH 234
            HLGI  D  +A+   EID+I+ +HTHHL E+G+++NG  LA+A KYG +VG ++IT D+
Sbjct: 179 HLGILDDQAVAEAVPEIDVILESHTHHLLEDGQVVNGVLLASAEKYGHYVGCVEITVDS- 237

Query: 235 TLKDILISTCDTKQLTGYPSDSDWLRRLSQKVKNSLEKKV 274
              + I   T  + +  +  +S  +    + +   E+K+
Sbjct: 238 VQRSINSKTASVQNMAEWTGESAETKAFLNEKEREAEEKL 277
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 46

A DNA sequence (GBSx0045) was identified in *S.agalactiae* <SEQ ID 145> which encodes the amino acid sequence <SEQ ID 146>. This protein is predicted to be UDP-sugar hydrolase. Analysis of this protein sequence reveals the following:

```
Possible site: 44

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.48   Transmembrane    5 -  21 (   5 -  21)

----- Final Results -----
               bacterial membrane --- Certainty=0.1192(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9605> which encodes amino acid sequence <SEQ ID 9606> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB15227 GB:Z99120 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 29/137 (21%), Positives = 71/137 (51%), Gaps = 13/137 (9%)

Query: 3   AMLFYAGADVAIINSGLIVQPFEKD-FSRKNLHESLPHQMRLAKLTVSSQELLEIYETIY 61
           A+  +   D++++NSG+I+ P +    ++ +LH   PH +    + ++ +EL E    ++
Sbjct: 305 ALKEWCETDISMVNSGVILGPLKAGPVTKLDLHRICPHPINPVAVRLTGEELKETI--VH 362

Query: 62  QQGQFLAQQKIHGMGFRGKCFGEVLHSGFDYKN----------GKIVYNEKDIDAKEEVI 111
              + + Q +I G+GFRG+  G+++++G + +             +I  N +DI+  ++
Sbjct: 363 AASEQMEQLRIKGLGFRGEVMGKMVYAGVEVETKRLDDGITHVTRITLNGEDIEKHKQYS 422

Query: 112 LVIVDQYYFASYFECLK 128
           + ++D +     F  ++
Sbjct: 423 VAVLDMFTLGKLFPLIR 439
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 47

A DNA sequence (GBSx0046) was identified in *S.agalactiae* <SEQ ID 147> which encodes the amino acid sequence <SEQ ID 148>. This protein is predicted to be unnamed protein product. Analysis of this protein sequence reveals the following:

```
Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3567(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein differs from AX026665 at the C-terminus:

```
Query: 181 SAKQHFVIRKK 191
           SAKQH +  +K
Sbjct: 181 SAKQHLLFVRK 191
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 149> which encodes the amino acid sequence <SEQ ID 150>. Analysis of this protein sequence reveals the following:

```
Possible site: 37

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3974(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 110/205 (53%), Positives = 147/205 (71%), Gaps = 15/205 (7%)
```

```
Query: 1   MRKEVTPEMLNYNKYPGPQFIHFENIVKSDDIEFQLVINEKSAFDVTVFGQRFSEILLKY 60
           M+KE++PEM NYNK+PGP+FIHFE  VK++ I+  L+ + K+AFD T FGQR++E+LLKY
Sbjct: 9   MKKEISPEMYNYNKFPGPKFIHFEEQVKAEGIDLLLLEDVKNAFDTTSFGQRYTEVLLKY 68

Query: 61  DFIVGDWGNEQLRLRGFYKDASTIRKNSRISRLEDYIKEYCNFGCAYFVLENPNPRDIKF 120
           D+IVGDWGNEQLRL+GFYKD+  I+K +RISRLEDYIKE+CNFGCAYFVLEN +P+DIKF
Sbjct: 69  DYIVGDWGNEQLRLKGFYKDSDDIKKTNRISRLEDYIKEFCNFGCAYFVLENLHPQDIKF 128

Query: 121 DDERPHKRRKS------RSKSQSSKSQTRNNRSQSNA--------NAHFTSKKRKDTKRR 166
           ++ER  +R+KS      R K   S  Q    +S+S          N  FTS+KR+   +
Sbjct: 129 EEERQPRRKKSPKSKSNRRKPNYSNQQPATPKSKSKRASKEKQPENQAFTSQKRRSNTKH 188

Query: 167 QERHIKEEQDKEMTSAKQHFVIRKK 191
           +E+  K  Q  ++ +   HF+IRKK
Sbjct: 189 KEKS-KRNQTSQLNTKISHFIIRKK 212
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 48

A DNA sequence (GBSx0047) was identified in *S.agalactiae* <SEQ ID 151> which encodes the amino acid sequence <SEQ ID 152>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3627(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9607> which encodes amino acid sequence <SEQ ID 9608> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB06225 GB:AP001515 unknown conserved protein [Bacillus halodurans]
 Identities = 205/349 (58%), Positives = 258/349 (73%), Gaps = 5/349 (1%)

Query: 18  PSIYSLTRDELIAWAIEHGEKKFRASQIWDWLYKKRVQSFDEMTNISKDFIALLNENFVV 77
           PSIY+L  +EL  W  E GE KFRA+QI++WLY+KRV+ F EMTN+SKD  A L ++F +
Sbjct: 17  PSIYTLQFEELEMWLKEQGEPKFRATQIFEWLYEKRVKQFQEMTNLSKDLRAKLEKHFNL 76

Query: 78  NPLKQRIVQESADGTVKYLFELPDGMLIETVLMRQHYGLSVCVTTQVGCNIGCTFCASGL 137
             LK    Q+S+DGT+K+LFEL DG  IETV+MR +YG SVCVTTQVGC +GCTFCAS L
Sbjct: 77  TTLKTVTKQQSSDGTIKFLFELHDGYSIETVVMRHNYGNSVCVTTQVGCRLGCTFCASTL 136

Query: 138 IKKQRDLNNGEITAQIMLVQKYFDERGQGERVSHIVVMGIGEPFDNYTNVLKFLRTVNDD 197
             +R+L  GEI AQ++  Q+  DE  QGERV  IVVMGIGEPFDNY  ++ FL+TVN D
Sbjct: 137 GGLKRNLEAGEIVAQVVEAQRAMDE--QGERVGSIVVMGIGEPFDNYQALMPFLKTVNHD 194

Query: 198 NGLAIGARHITVSTSGLAHKIREFANEGVQVNLAVSLHAPNNDLRSSIMRINRSFPLEKL 257
            GL IGARHITVSTSG+  KI +FA+EG+Q+N A+SLHAPN +LRS +M +NR++PL KL
Sbjct: 195 KGLNIGARHITVSTSGVVPKIYQFADEGLQINFAISLHAPNTELRSKLMPVNRAWPLPKL 254

Query: 258 FAAIEYYIETTNRRVTFEYIMLNGVNDTPENAQELADLTKKIRKLSYVNLIPYNPVSEHD 317
             AI YYI+ T RRVTFEY +  G ND  E+A+ELADL K I+   +VNLIP N V E D
Sbjct: 255 MDAIRYYIDKTGRRVTFEYGLFGGENDQVEHAEELADLIKDIK--CHVNLIPVNYVPERD 312

Query: 318 QYSRSPKERVEAFYDVLKKNGVNCVVRQEHGTDIDAACGQLRSNTMKRD 366
           Y R+P++++ AF   LK+ GVN  +R+E G DIDAACGQLR+   K +
Sbjct: 313 -YVRTPRDQIFAFERTLKERGVNVTIRREQGHDIDAACGQLRAKERKEE 360
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 153> which encodes the amino acid sequence <SEQ ID 154>. Analysis of this protein sequence reveals the following:

```
Possible site: 17

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2320(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 316/353 (89%), Positives = 339/353 (95%)

Query: 17  KPSIYSLTRDELIAWAIEHGEKKFRASQIWDWLYKKRVQSFDEMTNISKDFIALLNENFV 76
           KPSIYSLTRDELIAWA+E G+K+FRA+QIWDWLYKKRVQSF+EMTNISKDF+++LN++F
Sbjct: 2   KPSIYSLTRDELIAWAVERGQKQFRATQIWDWLYKKRVQSFEEMTNISKDFVSILNDSFC 61

Query: 77  VNPLKQRIVQESADGTVKYLFELPDGMLIETVLMRQHYGLSVCVTTQVGCNIGCTFCASG 136
           VNPLKQR+VQESADGTVKYLFELPDGMLIETVLMRQHYG SVCVTTQVGCNIGCTFCASG
Sbjct: 62  VNPLKQRVVQESADGTVKYLFELPDGMLIETVLMRQHYGHSVCVTTQVGCNIGCTFCASG 121

Query: 137 LIKKQRDLNNGEITAQIMLVQKYFDERGQGERVSHIVVMGIGEPFDNYTNVLKFLRTVND 196
           LIKKQRDLN+GEITAQIMLVQKYFD+R QGERVSH+VVMGIGEPFDNY NV+ FLR +ND
Sbjct: 122 LIKKQRDLNSGEITAQIMLVQKYFDDRKQGERVSHVVVMGIGEPFDNYKNVMCFLRVIND 181

Query: 197 DNGLAIGARHITVSTSGLAHKIREFANEGVQVNLAVSLHAPNNDLRSSIMRINRSFPLEK 256
           DNGLAIGARHITVSTSGLAHKIR+FANEGVQVNLAVSLHAPNNDLRSSIMR+NRSFPLEK
Sbjct: 182 DNGLAIGARHITVSTSGLAHKIRDFANEGVQVNLAVSLHAPNNDLRSSIMRVNRSFPLEK 241

Query: 257 LFAAIEYYIETTNRRVTFEYIMLNGVNDTPENAQELADLTKKIRKLSYVNLIPYNPVSEH 316
           LF+AIEYYIE TNRRVTFEYIMLN VND+ + AQELADLTK IRKLSYVNLIPYNPVSEH
Sbjct: 242 LFSAIEYYIEKTNRRVTFEYIMLNEVNDSIKQAQELADLTKTIRKLSYVNLIPYNPVSEH 301

Query: 317 DQYSRSPKERVEAFYDVLKKNGVNCVVRQEHGTDIDAACGQLRSNTMKRDRQK 369
           DQYSRSPKERV AFYDVLKKNGVNCVVRQEHGTDIDAACGQLRS TMK+DR+K
Sbjct: 302 DQYSRSPKERVLAFYDVLKKNGVNCVVRQEHGTDIDAACGQLRSKTMKKDREK 354
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 49

A DNA sequence (GBSx0048) was identified in *S.agalactiae* <SEQ ID 155> which encodes the amino acid sequence <SEQ ID 156>. This protein is predicted to be VanZF. Analysis of this protein sequence reveals the following:

```
Possible site: 47

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.61   Transmembrane   86 - 102 (  77 - 106)
   INTEGRAL    Likelihood = -8.60   Transmembrane   19 -  35 (  15 -  42)
   INTEGRAL    Likelihood = -5.15   Transmembrane  113 - 129 ( 109 - 134)

----- Final Results -----
               bacterial membrane --- Certainty=0.4843(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF36806 GB:AF155139 VanZF [Paenibacillus popilliae]
```

```
 Identities = 45/154 (29%), Positives = 68/154 (43%), Gaps = 36/154 (23%)

Query: 17  RRFVWMLVIIYCLIIVRMCFGPQIMIEGVSTPNVQRFGRIVAL-------LVPFNSFRSL 69
           R F+W+ V ++ L +V M  G           NV   GR   L       L+PF+S
Sbjct: 36  RHFLWVYVFLFYLALVYMMTG---------IGNVWVVGRYETLIRVSEINLLPFSS---- 82

Query: 70  DQLTSFKEIFWVIGQNVVNILLLFPLIIGLLSLKPSLRKYKSVILLAFLMSIFIECTQVV 129
           + +T++          ++NI+L  PL   L ++ P  R  K+     F  S+ IE TQ++
Sbjct: 83  EGVTTY----------ILNIILFMPLGFLLPTIWPQFRTIKNTACTGFFFSLAIELTQLL 132

Query: 130 LDILIDANRVFEIDDLWTNTLGGPFALWTYRNIK 163
                +R+ +IDDL  NTLG       YR  K
Sbjct: 133 ------NHRITDIDDLLMNTLGAIIGYLLYRAFK 160
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 50

A DNA sequence (GBSx0049) was identified in *S.agalactiae* <SEQ ID 157> which encodes the amino acid sequence <SEQ ID 158>. This protein is predicted to be multidrug resistance-like ATP-binding protein mdl. Analysis of this protein sequence reveals the following:

```
Possible site: 30

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -6.79   Transmembrane   18 -  34 (  17 -  36)
   INTEGRAL    Likelihood = -5.15   Transmembrane  247 - 263 ( 242 - 268)
   INTEGRAL    Likelihood = -2.81   Transmembrane  160 - 176 ( 158 - 176)
   INTEGRAL    Likelihood = -2.71   Transmembrane  141 - 157 ( 134 - 158)
   INTEGRAL    Likelihood = -1.12   Transmembrane   56 -  72 (  56 -  73)
   INTEGRAL    Likelihood = -0.69   Transmembrane  278 - 294 ( 277 - 294)

----- Final Results -----
               bacterial membrane --- Certainty=0.3718(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB06055 ABC transporter (ATP-binding protein) [Bacillus halodurans]
 Identities = 284/575 (49%), Positives = 406/575 (70%), Gaps = 2/575 (0%)

Query: 1   MSIIKNLWWFFKEEKKRYLIGILSLSLVAVLNLIPPKIMGSVIDAITTGKLTRPQLLWNL 60
           M +  +LWWFFK+EKK Y  GI+ L++V++L L+PP+++G ++D I  G LT P LL  +
Sbjct: 1   MKVFVDLWWFFKQEKKSYGFGIVMLAIVSLLTLVPPRVVGIIVDHIYEGTLTMPVLLQWI 60

Query: 61  LGLVLSALAMYGLRYIWRMYILGTSYKLGQVVRYRLFEHFTKMSPSFYQKYRTGDLMAHA 120
             L   AL +Y  RY+WR+ I G S +L +++R +L+ HFT M+  FYQK+RTGDLMAHA
Sbjct: 61  GVLAALALIVYVARYLWRVMIFGASLRLARLLRNQLYTHFTNMAAPFYQKHRTGDLMAHA 120

Query: 121 TNDINSLTRLAGGGVMSAVDASITALVTLITMFFTISWQMTLIAVIPLPLMALATSKLGR 180
           TNDI ++   AG GV++ VD+       ++TM  TISW++TLI+++P+PLMAL TS  G
Sbjct: 121 TNDIRAIQATAGQGVLTLVDSLTMGGFVILTMAITISWELTLISLLPMPLMALLTSYYGS 180

Query: 181 KTHETFKESQAAFSELNNKVQESVSGVKVTKSFGYQEQEIASFQEVNQMTFVKNMRTMTY 240
             H+ F  +QAAFS LN+KVQESV+GV+VTK+FG +EQ+I +F++ +     KN+
Sbjct: 181 LLHKRFHHAQAAFSSLNDKVQESVTGVRVTKAFGQEEQDIEAFRKQSDDVVKKNVAVARV 240

Query: 241 DVMFDPLVLLFIGASYVLTLAMGAFMISKGQVTVGDLVTFVTYLDMLVWPLMAIGFLFNM 300
           D +FDP + L +G SY L +  GA  +   Q+T+G L +F  YL +L+WP++A GFLFN+
Sbjct: 241 DALFDPTISLIVGLSYFLAIVFGARFVIAEQLTIGQLTSFTIYLGLLIWPMLAFGFLFNI 300
```

```
Query: 301 VQRGSVSYNRINSLLEQESDITDPLNPIRPVVNGTLRYDIDFFRYDN--EETLADIHFTL 358
           V+RG  SYNR++ LL+ + +ITD    I     G +   ID F Y N  E  LAD+ F L
Sbjct: 301 VERGRASYNRVSQLLQAKQEITDSRARIHVPPTGHVDVAIDQFVYPNQKEPALADVQFEL 360

Query: 359 EKGQTLGLVGQTGSGKTSLIKLLLREHDVTQGKITLNKHDIRDYRLSELRQLIGYVPQDQ 418
            +G+TLG+VG+TG+GKT+L++LL RE+D+ QG I L+   I  Y L  L+   G VPQD
Sbjct: 361 SEGETLGIVGKTGAGKTTLLRLLQREYDIKQGTIILDGRPIEHYTLDALKAAFGTVPQDH 420

Query: 419 FLFATSILENVRFGNPTLSINAVKKATKLAHVYDDIKQMPAGFETLIGEKGVSLSGGQKQ 478
           FLF+ +I +N+ F  P  +I+ + + ++LAH++DDI Q   G++T++GE+GV+LSGGQKQ
Sbjct: 421 FLFSATIADNIAFAKPDATISEIIQVSQLAHIHDDIIQFEQGYDTVVGERGVTLSGGQKQ 480

Query: 479 RIAMSRAMILDPDILILDDSLSAVDAKTEHAIIENLKTNRQGKSTIISAHRLSAVVHADL 538
           R++++RA++ +P+ILILDDSLSAVDAKTE AI+ +L+  R+GK+TII+AHRLSA+ HAD
Sbjct: 481 RVSIARALLANPNILILDDSLSAVDAKTEEAILSSLRAERKGKTTIITAHRLSAIKHADH 540

Query: 539 ILVMQDGRVIERGQHQELLNKGGWYAETYASQQLE 573
           ILVM DGR++ERG H+ L+  GGWY   Y  QQLE
Sbjct: 541 ILVMDDGRIVERGTHETLMEAGGWYRNMYERQQLE 575
```

There is also homology to SEQ ID 8.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 159> which encodes the amino acid sequence <SEQ ID 160>. Analysis of this protein sequence reveals the following:

```
Possible site: 23

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.75   Transmembrane  176 - 192 ( 173 - 197)
   INTEGRAL    Likelihood = -4.78   Transmembrane  267 - 283 ( 265 - 285)
   INTEGRAL    Likelihood = -4.09   Transmembrane   18 -  34 (  15 -  40)
   INTEGRAL    Likelihood = -2.13   Transmembrane  151 - 167 ( 150 - 169)
   INTEGRAL    Likelihood = -0.69   Transmembrane   85 - 101 (  85 - 101)

----- Final Results -----
               bacterial membrane --- Certainty=0.4100(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 172/609 (28%), Positives = 315/609 (51%), Gaps = 58/609 (9%)

Query: 1   MSIIKNLWWFFKEEKKRYLIGILSLSLVAVLNLIPPKIMGSVIDAITTGKLTRPQLLWNL 60
           M   +  W++FK  +  + +  +++ L   L +  P  +G  +  +  GK+ +   +  +
Sbjct: 2   MKTARFFWFYFKRYRFSFTVIAVAVILATYLQVKAPVFLGESLTEL--GKIGQAYYVAKM 59

Query: 61  LGLV-----LSAL--AMYGLRYIWRMYILGT---SYKLGQVV-------RYRLFEHFTKM 103
            G       LSA    M+ L   +   +L     S+ L +VV      R  LF    ++
Sbjct: 60  SGQTHFSPDLSAFNAVMFKLLMTYFFTVLANLIYSFLLTRVVSHSTNRMRKGLFGKLERL 119

Query: 104 SPSFYQKYRTGDLMAHATNDINSLTRLAGGGVMSAVDASITALVTLITMFFTISWQM--- 160
           + +F+ +++ G++++  T+D++         + ++++ S+  +VT I ++  + W M
Sbjct: 120 TVAFFDRHKDGEILSRFTSDLDN--------IQNSLNQSLIQVVTNIALYIGLVWMMFRQ 171

Query: 161 ------TLIAVIPLPLMALATS-KLGRKTHETFKESQAAFSELNNKVQESVSGVKVTKSF 213
                 IA  P+ L+ L  + +L RK         Q   S LN  + E++SG K
Sbjct: 172 DSRLALLTIASTPVALIFLVINIRLARKYTNI---QQQEVSALNAFMDETISGQKAIIVQ 228

Query: 214 GYQEQEIASF----QEVNQMTFVKNMRT------MTYDVMFDPLVLLFIGASYVLT-LAM 262
           G QE  + +F    + V Q TF + + +      M    + +  +++F+G++ VL+  +M
Sbjct: 229 GVQEDTMTAFLKHNERVRQATFKRRLFSGQLFPVMNGMSLINTAIVIFVGSTIVLSDKSM 288

Query: 263 GAFMISKGQVTVGDLVTFVTYLDMLVWPLMAIGFLFNMVQRGSVSYNRINSLLEQESDIT 322
            A         +G +VTFV Y      P+M I   +  +Q      +RI  + ++  ++
Sbjct: 289 PA------AAALGLVVTFVQYSQQYYQPMMQIASSWGELQLAFTGAHRIQEMFDETEEVR 342
```

```
Query: 323 DPLNPIRPVVNGTLRYD-IDFFRYDNEETLADIHFTLEKGQTLGLVGQTGSGKTSLIKLL 381
              P    +   +  + +DF     ++ L+D+     KG+ + +VG TGSGKT+++ L+
Sbjct: 343 PQNAPAFTSLKEAVAINHVDFGYLPGQKVLSDVSIVAPKGKMIAVVGPTGSGKTTIMNLI 402

Query: 382 LREHDVTQGKITLNKHDIRDYRLSELRQLIGYVPQDQFLFATSILENVRFGNPTLSINAV 441
             R +DV  G IT +  DIRDY L  LRQ +G V Q+  LF+ +I +N+RFG+ T+S + V
Sbjct: 403 NRFYDVDAGSITFDGRDIRDYDLDSLRQKVGIVLQESVLFSGTITDNIRFGDQTISQDMV 462

Query: 442 KKATKLAHVYDDIKQMPAGFETLIGEKGVSLSGGQKQRIAMSRAMILDPDILILDDSLSA 501
           + A +  H++D I  +P G+ T + +     S GQKQ I+++R ++ DP++LILD++ S
Sbjct: 463 ETAARATHIHDFIMSLPKGYNTYVSDDDNVFSTGQKQLISIARTLLTDPEVLILDEATSN 522

Query: 502 VDAKTEHAIIENLKTNRQGKSTIISAHRLSAVVHADLILVMQDGRVIERGQHQELLNKGG 561
           VD  TE  I   ++    G+++ + AHRL  ++ AD I+V++DG+VIE+G H ELL++ G
Sbjct: 523 VDTVTESKIQRAMEAIVAGRTSFVIAHRLKTILNADHIIVLKDGKVIEQGNHHELLHQKG 582

Query: 562 WYAETYASQ 570
           +YAE Y +Q
Sbjct: 583 FYAELYHNQ 591
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 51

A DNA sequence (GBSx0050) was identified in *S.agalactiae* <SEQ ID 161> which encodes the amino acid sequence <SEQ ID 162>. This protein is predicted to be mdlB (ATP-bindingprot). Analysis of this protein sequence reveals the following:

```
Possible site: 39

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.65   Transmembrane  164 - 180 ( 155 - 183)
   INTEGRAL    Likelihood = -5.15   Transmembrane   25 -  41 (  21 -  46)
   INTEGRAL    Likelihood = -4.88   Transmembrane  143 - 159 ( 133 - 163)
   INTEGRAL    Likelihood = -1.49   Transmembrane  251 - 267 ( 251 - 270)
   INTEGRAL    Likelihood = -1.33   Transmembrane   61 -  77 (  61 -  77)

----- Final Results -----
               bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB06054 ABC transporter (ATP-binding protein) [Bacillus halodurans]
 Identities = 278/582 (47%), Positives = 398/582 (67%), Gaps = 6/582 (1%)

Query: 1   MMKSNQWQVFKRLISYLRPYKWFTVLALSLLLLTTVVKNIIPLIASHFIDHYLT-NVNQT 59
           +    Q  VFKRL+SY   YK   ++A  LL + T  + + P+I   FID YLT     T
Sbjct: 9   LSSKEQRTVFKRLLSYAAHYKGQLMVAFLLLFIATGAQLLGPIIVKIFIDDYLTPRYFPT 68

Query: 60  AVLILVG--YYSMYVLQTLIQYFGNLFFARVSYSIVRDIRRDAFANMERLGMSYFDRTPA 117
            VL L+G  Y  +++   +I Y+    F +V+ SIV+ +R D F++++RLG+S+FD+TPA
Sbjct: 69  DVLFLLGAGYLVLHLTAVIIDYYQLFLFQKVALSIVQRLRIDVFSSVQRLGLSFFDQTPA 128

Query: 118 GSIVSRITNDTEAISDMFSGILSSFISAIFIFTVTLYTMLMLDIKLTGLVALLLPVIFIL 177
           G +VSRITNDTE+I +++  +L++F+  I          M  L++ L     +LLP+IF L
Sbjct: 129 GGLVSRITNDTESIKELYVTVLATFVQNIIFLIGIFAAMFYLNVTLAIYCLVLLPLIFAL 188

Query: 178 VNVYRKKSVTVIAKTRSLLSDINSKLSESIEGIRIVQAFGQEERLKTEFEEINKEHVVYA 237
           + VYRK S    A     LS +N +++ESI+G+ I+Q F QE R++ EF  IN EH +
Sbjct: 189 MQVYRKYSSRFYADMSEKLSLLNGRINESIQGMAIIQMFRQERRMRKEFSAINDEHFLAG 248

Query: 238 NRSMALDSLFLRPAMSLLKLLAYAVLMAYFGFTGVKGGLTAGLMYAFIQYVNRLFDPLIE 297
```

```
             +SM LD L LRPA+ +L +LA  ++++YFG   +   +  G++YAF+ Y++R F+P+ +
Sbjct: 249 MKSMKLDGLLLRPAVDVLSILALMLILSYFGIMSMDTAVEIGVVYAFVNYLDRFFEPVNQ 308

Query: 298 VTQNFSTLQTSMVSAGRVFDLIDETGFEPSQKNTE--AFVREGNIEFKNVSFSYDGKKQI 355
           +    S .Q ++VSAGRVF L+D     P ++  E  A + EGN+EF+NVSFSYDGK  +
Sbjct: 309 MMMRLSMFQQAIVSAGRVFKLMDHRELAPDREGNEHPAIIGEGNVEFRNVSFSYDGKTNV 368

Query: 356 LDNVSFSVKKGETIAFVGATGSGKSSIINVFMRFYEFQSGQVLLDGKDIRDYSQEQLRKN 415
           L N+SF+VKKGET+A VG TGSGK+SIINV MRFY  Q G++L+DGK +  +   +LR
Sbjct: 369 LKNISFTVKKGETVALVGHTGSGKTSIINVLMRFYPLQDGEILIDGKPLTSFENNELRAK 428

Query: 416 IGLVLQDPFLYHGTIKSNIKMY-QDITDQEVQDAAEFVDADQFIQKLPDKYDAAVSERGS 474
           +GLVLQDPFLY GTI SNI++Y Q I+D  ++ AA FV AD FI++L   Y+  V+ERG+
Sbjct: 429 VGLVLQDPFLYTGTIASNIRLYDQAISDDRIKRAASFVRADGFIERLSHGYETKVTERGA 488

Query: 475 SFSTGQRQLLAFARTVASKPKILILDEATANIDSETEQIVQDSLAKMRQGRTTIAIAHRL 534
           +FS+GQRQLL+FART+  +P ILILDEATA++D+ETE+ +Q++L +M+QGRTTIAIAHRL
Sbjct: 489 TFSSGQRQLLSFARTMVREPAILILDEATASVDTETEEAIQEALERMKQGRTTIAIAHRL 548

Query: 535 STIQDANCIYVLDRGKIIESGNHESLLDLKGTYYRMYQLQAG 576
           STI+DA+ I VL +G+I+E G H+ L+  KG Y +MY LQ G
Sbjct: 549 STIKDADQILVLHQGEIVERGTHDELIAKKGLYQKMYVLQKG 590
```

There is also homology to SEQ ID 160.

A related GBS gene <SEQ ID 8481> and protein <SEQ ID 8482> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:     -4.63
GvH: Signal Score (-7.5): -5.85
     Possible site: 39
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 5 value:  -8.65 threshold:  0.0
   INTEGRAL    Likelihood = -8.65   Transmembrane  164 - 180 ( 155 - 183)
   INTEGRAL    Likelihood = -5.15   Transmembrane   25 -  41 (  21 -  46)
   INTEGRAL    Likelihood = -4.88   Transmembrane  143 - 159 ( 133 - 163)
   INTEGRAL    Likelihood = -1.49   Transmembrane  251 - 267 ( 251 - 270)
   INTEGRAL    Likelihood = -1.33   Transmembrane   61 -  77 (  61 -  77)
   PERIPHERAL  Likelihood =  3.02     483
 modified ALOM score:   2.23

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01277(322 - 2028 of 2340)
EGAD|108578|BS0971(2 - 667 of 673) hypothetical protein {Bacillus subtilis} OMNI|NT01BS1137
conserved hypothetical protein GP|2226165|emb|CAA74449.1||Y14080 hypothetical protein
{Bacillus subtilis} GP|2633307|emb|CAB12811.1||Z99109 similar to ABC transporter (ATP-
binding protein) {Bacillus subtilis} PIR|H69828|H69828 ABC transporter (ATP-binding
protein) homolog yheH - Bacillus subtilis
%Match = 28.5
%Identity = 40.8  %Similarity = 69.1
Matches = 234  Mismatches = 171  Conservative Sub.s = 162

162       192       222       252       282       312       342       372
RLLFQHIDYQLLCTQTLS*LCKTAESSSEVSIKSC*IKVVGMLKRMPHSN*KWRKHLMKSNQWQVFKRLISYLRPYKWFT
                                                          :: | |  |   |: :
                                                        MKIGKTLWRYALLYRKLL
                                                                  10
```

```
402       432       462                                                480
VLALSLLLLTTVVKNIIPLIASHFIDHYLTNVNQT---------------------~~~-------------------A
: |: || :    :    |:|    || ::  : :|
ITAVLLLTVAVGAELTGPFIGKKMIDDHILGIEKTWYEAAEKDKNAVQFHGVSYV~~~~AAEKLTKQELFQFYQPEIKGM
          30        40        50        60        70                  140

510       540       570       600       630       660       690       720
VLILVGYYSMYVLQTLIQYFGNLFFARVSYSIVRDIRRDAFANMERLGMSYFDRTPAGSIVSRITNDTEAISDMFSGILS
||::  |  : |:  : ||  : ::   :  |:: :|:| |:::::: : |||  ||| :|:||||||||| |::  :||
VLLICLYGGLLVFSVFFQYGQHYLLQMSANRIIQKMRQDVFSHIQKMPIRYFDNLPAGKVVARITNDTEAIRDLYVTVLS
         160       170       180       190       200       210       220

750       777       807       837       867       897       927       957
SFISAIFIFTVTLYTML-MLDIKLTGLVALLLPVIFILVNVYRKKSVTVIAKTRSLLSDINSKLSESIEGIRIVQAFGQE
:|:::  |:   ::| | :||:||  :   ::|:|::   :||: :        | ||: ||||:|::|||:|: |:||| ::
TFVTS-GIYMFGIFTALFLLDVKLAFVCLAIVPIIWLWSVIYRRYASYYNQKIRSINSDINAKMNESIQGMTIIQAFRHQ
        240       250       260       270       280       290       300

987      1017      1047      1077      1107      1131      1161      1191
ERLKTEFEEINKEHVVYANRSMALDSLFLRPAMSLLKLLAYAVLMAYFGFTGVK--GGLTAGLMYAFIQYVNRLFDPLIE
:     ||||:|: |  : || : |:||      ::::: ||:  |: :||    :    | :: |::|||: |:|||| |:
KETMREFEELNESHFYFQNRMLNLNSLMSHNLVNVIRNLAFVCLIWHFGGASLNAAGIVSIGVLYAFVDYLNRLFQPITG
        320       330       340       350       360       370       380

1221      1251      1281      1311      1341      1371      1401      1431
VTQNFSTLQTSMVSAGRVFDLIDETGFEPSQKNTEAFVREGNIEFKNVSFSYDGKKQILDNVSFSVKKGETIAFVGATGS
:    || |: : |||||||:|::|    | : :  :     | :||::|||:|   :::| ::||: :||||:|:|| |||
IVNQFSKLELARVSAGRVFELLEEKNTEEAGEPAKERAL-GRVEFRDVSFAYQEGEEVLKHISFTAQKGETVALVGHTGS
        400       410       420       430       440       450       460

1461      1491      1521      1551      1581      1611      1638      1668
GKSSIINVFMRFYEFQSGQVLLDGKDIRDYSQEQLRKNIGLVLQDPFLYHGTIKSNIKMYQD-ITDQEVQDAAEFVDADQ
|||||:|:: |||: | | ||:||| | : |:::|| ::|:||||||:|: ||| ||: :   : :|::|:::|    | |:
GKSSILNLLFRFYDAQKGDVLIDGKSIYNMSRQELRSHMGIVLQDPYLFSGTIGSNVSLDDERMTEEEIKNALRQVGAEP
         480       490       500       510       520       530       540

1698      1728      1758      1788      1818      1848      1878      1908
FIQKLPDKYDAAVSERGSSFSTGQRQLLAFARTVASKPKILILDEATANIDSETEQIVQDSLAKMRQGRTTIAIAHRLST
:::|||    :   | |:||::|:|:|||::||| :|   | |||||||||:||:||| ::| :|  ::|||||  |||||||
LLKKLPKGINEPVIEKGSTLSSGERQLISFARALAFDPAILILDEATAHIDTETEAVIQKALDVVKQGRTTFVIAHRLST
         560       570       580       590       600       610       620

1938      1968      1998      2028      2058      2088      2118      2148
IQDANCIYVLDRGKIIESGNHESLLDLKGTYYRMYQLQAGMMEV*KI*TIQKA*SVRFRGWSSYSSKPFLYFTISV**GQ
|::|: | |||:|:|:| |||| |: |:| ||:||:|| |
IRNADQILVLDKGEIVERGNHEELMALEGQYYQMYELQKGQKHSIA
         640       650       660       670
```

There is also homology to SEQ IDs 330, 4634 and 5788.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 52

A DNA sequence (GBSx0051) was identified in *S.agalactiae* <SEQ ID 163> which encodes the amino acid sequence <SEQ ID 164>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0635(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9609> which encodes amino acid sequence <SEQ ID 9610> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA25224 GB:M87483 anthranilate synthase beta subunit
           [Lactococcus lactis]
 Identities = 101/191 (52%), Positives = 133/191 (68%), Gaps = 4/191 (2%)

Query: 14  MLLLVDNYDSFTYNLKQYLSVYKEVFVIKNDVPNLFLLAESAEAIVLSPGPGHPKDAGKM 73
           M+L++DNYDSFTYNL QY+ V  +V V+KND  +L  +AE A+A++ SPGPG P DAGKM
Sbjct: 1   MILIIDNYDSFTYNLVQYVGVLTDVAVVKNDDDSLGNMAEKADALIFSPGPGWPADAGKM 60

Query: 74  VELINQFIGKKPILGICLGHQALAECLGGRLNLANHVMHGKQSWVTINDHTSLFKGIDSP 133
             LI QF G+KPILGICLG QA+ E  GG+L LA+ VMHGK S V       +F  + S
Sbjct: 61  ETLIQQFAGQKPILGICLGFQAIVEVFGGKLRLAHQVMHGKNSQVRQTSGNLIFNHLPSK 120

Query: 134 TQVMRYHSLVVTD---LPENIAVIARSNEDNEIMAFHCPSLKVYAMQFHPESIGSIDGMK 190
             VMRYHS+V+ +   LP+  A+ A + +D EIMA      ++Y +QFHPESIG++DGM
Sbjct: 121 FLVMRYHSIVMDEAVALPD-FAITAVATDDGEIMAIENEKEQIYGLQFHPESIGTLDGMT 179

Query: 191 MIENFLTLIND 201
           MIENF+  +N+
Sbjct: 180 MIENFVNQVNE 190
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 165> which encodes the amino acid sequence <SEQ ID 166>. Analysis of this protein sequence reveals the following:

```
Possible site: 57

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3183(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 104/186 (55%), Positives = 131/186 (69%)

Query: 14  MLLLVDNYDSFTYNLKQYLSVYKEVFVIKNDVPNLFLLAESAEAIVLSPGPGHPKDAGKM 73
           M+LL+DNYDSFTYNL QYLS + E  V+ N  PNL+ +A+ A A+VLSPGPG PK+A +M
Sbjct: 1   MILLIDNYDSFTYNLAQYLSEFDETIVLYNQDPNLYDMAKKANALVLSPGPGWPKEANQM 60

Query: 74  VELINQFIGKKPILGICLGHQALAECLGGRLNLANHVMHGKQSWVTINDHTSLFKGIDSP 133
            +LI  F   KPILG+CLGHQA+AE LGG L LA  VMHG+QS +       SLF+ +
Sbjct: 61  PKLIQDFYQTKPILGVCLGHQAIAETLGGTLRLAKRVMHGRQSTIETQGPASLFRSLPQE 120

Query: 134 TQVMRYHSLVVTDLPENIAVIARSNEDNEIMAFHCPSLKVYAMQFHPESIGSIDGMKMIE 193
             VMRYHS+VV  LP+  +V AR  +D EIMAF   +L ++ +QFHPESIG+ DGM MI
Sbjct: 121 ITVMRYHSIVVDQLPKGFSVTARDCDDQEIMAFEHHTLPLFGLQFHPESIGTPDGMTMIA 180

Query: 194 NFLTLI 199
           NF+  I
Sbjct: 181 NFIAAI 186
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 53

A DNA sequence (GBSx0052) was identified in *S.agalactiae* <SEQ ID 167> which encodes the amino acid sequence <SEQ ID 168>. Analysis of this protein sequence reveals the following:

```
Possible site: 58

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -8.17   Transmembrane  117 - 133 ( 108 - 140)
   INTEGRAL    Likelihood = -1.70   Transmembrane  150 - 166 ( 150 - 166)

----- Final Results -----
               bacterial membrane --- Certainty=0.4270(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB12877 GB:Z99109 similar to biotin biosynthesis [Bacillus subtilis]
 Identities = 70/168 (41%), Positives = 106/168 (62%)

Query: 8   YIALMVALLIVLGFIPGIPLGFIPVPIVLQNLGVMLAGALLGSRKGFLAVAIFLLLVAIG 67
           +IA+  AL+ VLGF+P + L F PVPI LQ LGVMLAG++L  +  FL+  +FLLLVA G
Sbjct: 9   HIAIFTALMAVLGFMPPLFLSFTPVPITLQTLGVMLAGSILRPKSAFLSQLVFLLLVAFG 68

Query: 68  APFLPGGRSGLVTLFGPTAGYLLTYPFAAFFIGLGLEKVKTTKLWVQFLIIWIFGVLLID 127
           AP LPGGR G    FGP+AG+L+ YP A++ I L   +++   +   F    +FG++ I
Sbjct: 69  APLLPGGRGGFGVFFGPSAGFLIAYPLASWLISLAANRLRKVTVLRLFFTHIVFGIIFIY 128

Query: 128 ICGSIVLSFQTSLPLTKSLFSNLIFIPGDTLKASICLIIYRKFANRLT 175
           + G  V +F   + L+++ F +L ++PGD +KA++  +  K    L+
Sbjct: 129 LLGIPVQAFIMHIDLSQAAFMSLAYVPGDLIKAAVSAFLAIKITQALS 176
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 169> which encodes the amino acid sequence <SEQ ID 170>. Analysis of this protein sequence reveals the following:

```
Possible site: 51

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.03   Transmembrane  113 - 129 ( 109 - 139)
   INTEGRAL    Likelihood = -8.97   Transmembrane   55 -  71 (  52 -  76)
   INTEGRAL    Likelihood = -7.54   Transmembrane   10 -  26 (   6 -  38)
   INTEGRAL    Likelihood = -5.79   Transmembrane   86 - 102 (  81 - 105)
   INTEGRAL    Likelihood = -2.87   Transmembrane   33 -  49 (  28 -  51)
   INTEGRAL    Likelihood = -1.97   Transmembrane  150 - 166 ( 150 - 168)

----- Final Results -----
               bacterial membrane --- Certainty=0.5012(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 80/168 (47%), Positives = 108/168 (63%), Gaps = 1/168 (0%)

Query: 3   TRTTTYIALMVALLIVLGFIPGIPLGFIPVPIVLQNLGVMLAGALLGSRKGFLAVAIFLL 62
           T+    +A+M  L+I+LGFIP IPLGFIPVPIVLQNLGVMLAG +LG +KG L+V +F L
Sbjct: 4   TKELVKVAMMTTLIIILGFIPAIPLGFIPVPIVLQNLGVMLAGLMLGGKKGTLSVFLF-L 62

Query: 63  LVAIGAPFLPGGRSGLVTLFGPTAGYLLTYPFAAFFIGLGLEKVKTTKLWVQFLIIWIFG 122
           ++ +  P   G R+ +  L GP+AGY++ Y       L      +    + FL + I G
Sbjct: 63  VIGLFLPVFSGSRTTIPVLMGPSAGYVIAYLLVPIVFSLLYRNWFSKSTPLAFLALLISG 122

Query: 123 VLLIDICGSIVLSFQTSLPLTKSLFSNLIFIPGDTLKASICLIIYRKF 170
           V+L+D+ G+I LS  T + L  SL SNL+FIPGDT+KA I  II  K+
Sbjct: 123 VVLVDVLGAIWLSAYTGMSLVTSLLSNLVFIPGDTIKAIIATIIAVKY 170
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 54

A DNA sequence (GBSx0053) was identified in *S.agalactiae* <SEQ ID 171> which encodes the amino acid sequence <SEQ ID 172>. Analysis of this protein sequence reveals the following:

```
Possible site: 17

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3914(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 55

A DNA sequence (GBSx0054) was identified in *S.agalactiae* <SEQ ID 173> which encodes the amino acid sequence <SEQ ID 174>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1864(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9611> which encodes amino acid sequence <SEQ ID 9612> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB05467 GB:AP001513 biotin synthase [Bacillus halodurans]
 Identities = 133/316 (42%), Positives = 201/316 (63%), Gaps = 2/316 (0%)

Query: 17  NYIHLADEILSGKTSISYEQALEILNS-DENWWEIYAAALYLKNQVSRNNIRLNVLLSAK 75
           N+I LA E++ GK  IS  +AL ILNS D+    +   A  ++        ++LN++++AK
Sbjct: 2   NWIQLAQEVIEGKR-ISENEALAILNSPDDELLLLLQGAFTIRQTYYGKKVKLNMIMNAK 60

Query: 76  QGLCAENCGYCSQSKESTADIDKFGLLPQNVILKQAIVAHQNGASVFCIAMSGTKPSKRE 135
            G C ENCGYCSQS  S A ID + ++ +  IL+ A  AH+     +CI  SG  P+ R+
Sbjct: 61  SGFCPENCGYCSQSSISKAPIDAYPMVNKETILEGAKRAHELNVGTYCIVASGRGPTNRD 120

Query: 136 IEQLCQVIPEIKKSLPLEICLTAGFLDREQLHQLKQAGIDRINHNLNTPEENYPNIATTH 195
           I+ + + + EIK +  L+IC   G L  EQ  QLK AG+DR NHN+NT   ++  I T+H
Sbjct: 121 IDHVTEAVREIKDTYGLKICACLGILKPEQAEQLKAAGVDRYNHNVNTSARHHDQITTSH 180

Query: 196 SFKDRCDTLERIHNEDIDVCSGFICGMGESDEGLITLAFRLKELDPYSIPVNFLLAVEGT 255
           +++DR +T+E + +  I  CSG I GM E+ E ++ +AF+L+ELD  SIPVNFL A++GT
Sbjct: 181 TYEDRVNTVEVVKHSGISPCSGVIVGMKETKEDVVDMAFQLRELDADSIPVNFLHAIDGT 240
```

```
Query: 256 PLGKYNYLTPIKCLKIMAMLRFVFPFKELRLSAGREVHFENFESLVTLLVDSTFLGNYLT 315
           PL   + LTPI CLK++++ R+V P KE+R+S GREV+ ++ + L     +S F+G+YLT
Sbjct: 241 PLQGVHELTPIYCLKVLSLFRYVCPTKEIRISGGREVNLKSLQPLGLYAANSIFIGDYLT 300

Query: 316 EGGRNQHTDIEFLEKL 331
             G+ +  D + L+ L
Sbjct: 301 TAGQEETADHQILKDL 316
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 56

A DNA sequence (GBSx0055) was identified in *S.agalactiae* <SEQ ID 175> which encodes the amino acid sequence <SEQ ID 176>. Analysis of this protein sequence reveals the following:

```
Possible site: 24

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3440(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9613> which encodes amino acid sequence <SEQ ID 9614> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 57

A DNA sequence (GBSx0056) was identified in *S.agalactiae* <SEQ ID 177> which encodes the amino acid sequence <SEQ ID 178>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1985(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 58

A DNA sequence (GBSx0057) was identified in *S.agalactiae* <SEQ ID 179> which encodes the amino acid sequence <SEQ ID 180>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.11   Transmembrane  347 - 363 ( 347 - 363)

----- Final Results -----
               bacterial membrane --- Certainty=0.1044(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAC11722 GB:AL445064 acetyl-CoA acetyltransferase related
          protein [Thermoplasma acidophilum]
 Identities = 113/388 (29%), Positives = 181/388 (46%), Gaps = 31/388 (7%)

Query: 4   RDVYIGFGLRTPIGIKGKQFKHYR-PELLGAHLLNQIKKIESESNID-----SIICGNTV 57
           RDV+I    RT IG  G+ F   + P+L GA     IK +  E+++D      +I GN +
Sbjct: 2   RDVFIVAAKRTAIGKFGRSFSKLKAPQLGGA----AIKAVMDEAHVDPASVEEVIMGNVI 57

Query: 58  --GTGGNIGRLMTLFSDYESYIPVQTIDMQCASSSSALFFGYLKISTGINEKVLVGGIES 115
             G G N            + +   T+++ CAS   A+     +I+ G  + V+ GG+ES
Sbjct: 58  QAGNGQNPAGQAAFHGGLPNSVLKYTVNVVCASGMLAVESAAREIALGERDLVIAGGMES 117

Query: 116 SSLQPMR-----RYAKEDNRNGEYTVAQ-FSPDSYAETVMLE----GAQRVCQKYGFRRE 165
            S  P       R+  +   + Y +      D   +    E     A+R  +K+G  RE
Sbjct: 118 MSNAPFLLPADLRWGPKHLLHKNYKIDDAMLTDGLLDAFYFEHMGVSAERTSRKFGITRE 177

Query: 166 MLDKLAFLSHKRALTAKQGGYLEEVILPMEGM-RDQGVRKLKETFFQKLPRLMENSPLLT 224
           M D+ +  S++RA+ A + G   + I+  EG+  D+G+RK       +LP   + + +LT
Sbjct: 178 MADEYSVQSYERAIRATESGEFADEIVQFEGLDHDEGIRKTTMEDLARLPPAFDKNGILT 237

Query: 225 IGNVCLMHDAAAFLTLQSQKT--EFRIVHIVEVAG------DPKLSPELVHTATEKLLTE 276
            GN   + D  + L + S+K   E+ +  I  + G      DP    E    AT KLL +
Sbjct: 238 AGNSAQLSDGGSALMIASEKAINEYGLKPIARITGYEQASLDPLDFVEAPIPATRKLLEK 297

Query: 277 THTKISDYDAIEWNEPFAAIDALFNHYYPEEREKFNIFGGTLAYGHPYACSGIINILHLM 336
            H  I  YD +E NE F+    +  +    + E+FN+ GG +A GHP   SG   I+ LM
Sbjct: 298 QHKSIDYYDLVEHNEAFSIASVIVRNELKIDNERFNVNGGAVAIGHPIGNSGARIIVTLM 357

Query: 337 QALKYKNKPMGLTAIAGAGGVGMAISIE 364
            ALK+++   GL  +  GG    +++E
Sbjct: 358 NALKHRHLKTGLATLCHGGGGAHTLTLE 385
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 181> which encodes the amino acid sequence <SEQ ID 182>. Analysis of this protein sequence reveals the following:

```
     Possible site: 22
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.28   Transmembrane  345 - 361 ( 345 - 361)

----- Final Results -----
               bacterial membrane --- Certainty=0.1510(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB03328 GB:AB035449 acetyl-CoA c-acetyltransferase
          [Staphylococcus aureus]
 Identities = 115/382 (30%), Positives = 184/382 (48%), Gaps = 29/382 (7%)
```

```
Query: 1   MTDVYIAAGLRTPIGLVGKQFAKEQPEILGAKLINALQNKYPV---PIDQVICGNTVGTG 57
           M    I A  RT  G  G       +PE L   L     + KYP      ID V+ GN VG G
Sbjct: 1   MNQAVIVAAKRTAFGKYGGTLKHLEPEQLLKPLFQHFKEKYPEVISKIDDVVLGNVVGNG 60

Query: 58  GNIGRLMTLYSHLGESVSALTVDMQCASAGAALSVGYAKIKAGMASNLLVGGIESSS--- 114
           GNI R   L + L +S+  +T+D QC S   ++       I+AG      + GG+ES+S
Sbjct: 61  GNIARKALLEAGLKDSIPGVTIDRQCGSGLESVQYACRMIQAGAGKVYIAGGVESTSRAP 120

Query: 115 ---LQPESVYASADWRQGAYKVAQFSPDSISPFAMIEGAERVAREHGFTKEYLNHWTLRS 171
              +P SVY +A       Y+ A F+P+    P +MI+GAE VA+ +  ++E  + +  RS
Sbjct: 121 WKIKRPHSVYETA--LPEFYERASFAPEMSDP-SMIQGAENVAKMYDVSRELQDEFAYRS 177

Query: 172 HQKASYCQEQALLADLILDLSGA-----SDQGIRPRLSSKVLSKVPPILGEGHVISAANA 226
           HQ  +   +   ++  IL ++       +D+ ++  +       +  P++ +G  ++AAN+
Sbjct: 178 HQLTAENVKNGNISQEILPITVKGEIFNTDESLKSHIPKDNFGRFKPVI-KGGTVTAANS 236

Query: 227 CLTHDAAAFLQLSSQPSAFKL--------IDVVEVAGDPQRSPLMVIKASQVLLEKHGLG 278
           C+ +D A  L +  +  A++L          D V V  D     +  + A   LL++ L
Sbjct: 237 CMKNDGAVLLLIMEKDMAYELGFEHGLLFKDGVTVGVDSNFPGIGPVPAISNLLKRNQLT 296

Query: 279 MADMTAIEWNEAFAVIDGLFETHYPDLLDRYNIFGGALAYGHPYGASAAIIILHLMRALE 338
           + ++  IE NEAF+      +        + NI+GGALA GHPYGAS A ++  L    +
Sbjct: 297 IENIEVIEINEAFSAQVVACQQALNISNTQLNIWGGALASGHPYGASGAQLVTRLFYMFD 356

Query: 339 IKNGRYGIAAIAAAGGQGFAVL 360
            +     IA++   GG G A L
Sbjct: 357 KET---MIASMGIGGGLGNAAL 375
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 182/362 (50%), Positives = 243/362 (66%), Gaps = 2/362 (0%)

Query: 5   DVYIGFGLRTPIGIKGKQFKHYRPELLGAHLLNQIKKIESESNIDSIICGNTVGTGGNIG 64
           DVYI  GLRTPIG+ GKQF   +PE+LGA L+N ++  +    ID +ICGNTVGTGGNIG
Sbjct: 3   DVYIAAGLRTPIGLVGKQFAKEQPEILGAKLINALQN-KYPVPIDQVICGNTVGTGGNIG 61

Query: 65  RLMTLFSDYESYIPVQTIDMQCASSSSALFFGYLKISTGINEKVLVGGIESSSLQPMRRY 124
           RLMTL+S     +   T+DMQCAS+ +AL  GY KI  G+   +LVGGIESSSLQP   Y
Sbjct: 62  RLMTLYSHLGESVSALTVDMQCASAGAALSVGYAKIKAGMASNLLVGGIESSSLQPESVY 121

Query: 125 AKEDNRNGEYTVAQFSPDSYAETVMLEGAQRVCQKYGFRREMLDKLAFLSHKRALTAKQG 184
           A  D R G Y VAQFSPDS +   M+EGA+RV +++GF +E L+     SH++A   ++
Sbjct: 122 ASADWRQGAYKVAQFSPDSISPFAMIEGAERVAREHGFTKEYLNHWTLRSHQKASYCQEQ 181

Query: 185 GYLEEVILPMEGMRDQGVR-KLKETFFQKLPRLMENSPLLTIGNVCLMHDAAAFLTLQSQ 243
             L ++IL + G  DQG+R +L      K+P ++    +++  N CL HDAAAFL L SQ
Sbjct: 182 ALLADLILDLSGASDQGIRPRLSSKVLSKVPPILGEGHVISAANACLTHDAAAFLQLSSQ 241

Query: 244 KTEFRIVHIVEVAGDPKLSPELVHTATEKLLTETHTKISDYDAIEWNEPFAAIDALFNHY 303
            + F+++ +VEVAGDP+ SP +V  A++ LL +    ++D  AIEWNE FA ID LF  +
Sbjct: 242 PSAFKLIDVVEVAGDPQRSPLMVIKASQVLLEKHGLGMADMTAIEWNEAFAVIDGLFETH 301

Query: 304 YPEEREKFNIFGGTLAYGHPYACSGIINILHLMQALKYKNKPMGLTAIAGAGGVGMAISIEY 365
           YP+  +++NIFGG LAYGHPY  S  I ILHLM+AL+ KN   G+ AIA AGG G A+ ++Y
Sbjct: 302 YPDLLDRYNIFGGALAYGHPYGASAAIIILHLMRALEIKNGRYGIAAIAAAGGQGFAVLLKY 363
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 59

A DNA sequence (GBSx0058) was identified in *S.agalactiae* <SEQ ID 183> which encodes the amino acid sequence <SEQ ID 184>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
```

```
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.82   Transmembrane  149 - 165 ( 148 - 165)

----- Final Results -----
               bacterial membrane --- Certainty=0.2529(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB12876 GB:Z99109 similar to long-chain fatty-acid-CoA ligase
           [Bacillus subtilis]
 Identities = 90/382 (23%), Positives = 158/382 (40%), Gaps = 24/382 (6%)

Query: 47  ISTHSLLNQLVRFVSKLCQKALPIICKPNLTHNEISRLEKEV--QYAPQLADFGVLSSGT 104
           IS   L+  L  F +KL     P++   N    +IS    +      P+   +   +SG+
Sbjct: 95  ISNADLVVTLAFFKNKLTDSQTPVVLLDNCMA-DISEAAADPLPTIDPEHPFYMGFTSGS 153

Query: 105 TADAKLLWRSFTSWSDFFSIQNAYFSVTSNSKLFIQGDFSFTGNLNLALSLLLLGGTLVV 164
           T   K   RS  SW + F+     FS++S+ K+ I G    +  L  A+S L LGGT+ +
Sbjct: 154 TGKPKAFTRSHRSWMESFTCTETDFSISSDDKVLIPGALMSSHFLYGAVSTLFLGGTVCL 213

Query: 165 TQKNSVKYWQTLWEKTGVTHLYLLPSYLKLVEQYSKETALDNKTIITSSQYVSDSLLEGL 224
            +K S    +    +  ++ LY +P+    + +         K I + + + ++S  + L
Sbjct: 214 LKKFSPAKAKEWLCRESISVLYTVPTMTDALARIEGFPDSPVKIISSGADWPAES-KKKL 272

Query: 225 YRKHPKVSVKIFYGASELNYVSWYDGRDIRDKPQYVGEIVPNVAVRIKE---------- 273
               P + +  FYG SEL++V++    D + KP   G    NV + I+
Sbjct: 273 AAAWPHLKLYDFYGTSELSFVTFSSPEDSKRKPHSAGRPFHNVRIEIRNAGGERCQPGEI 332

Query: 274 GRIFVKTPYSICG-----LSSEYCAGDYGELID--GKLYLFGRGGDWCNQSGIKLYLPRL 326
           G+IFVK+P    G        E+   D    +D  G LY+ GR        G+ ++   +
Sbjct: 333 GKIFVKSPMRFSGYVNGSTPDEWMTVDDMGYVDEEGFLYISGRENGMIVYGGLNIFPEEI 392

Query: 327 IEKIKTCPYIKDAVAFTKESQSHGQESHCCIVLIENQMQQECLKWLSEHFEKKYGFKHYH 386
              +  CP ++ A       +  G+ +    V++ N   +    W  +        K +
Sbjct: 393 ERVLLACPEVESAAVVGIPDEYWGEIA--VAVILGNANARTLKAWCKQKLASYKIPKKWV 450

Query: 387 IVSKIPLMPSGKIDYQQLKRQL 408
              +P   SGKI   ++K+ L
Sbjct: 451 FADSLPETSSGKIARSRVKKWL 472
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 185> which encodes the amino acid sequence <SEQ ID 186>. Analysis of this protein sequence reveals the following:

```
Possible site: 52

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2487(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 154/413 (37%), Positives = 235/413 (56%), Gaps = 9/413 (2%)

Query: 1   MLESLKTIVKTNSDKKLFDGD-LQVSYGEFYNLVR-QDMASQDNRKHVISTHSLLNQLVR 58
           ML  L+   K   +KK    D + ++Y E +  V  +D   +D+  ++IS    LNQL+
Sbjct: 1   MLTKLEYWAKQCPNKKAIVADQISLTYQELWQAVLIKDQTIKDSVPYIISHSRYLNQLLS 60

Query: 59  FVSKLCQKALPIICKPNLT---HNEISRLEKEVQYAPQLADFGVLSSGTTADAKLLWRSF 115
           F+  L + + PII  PN++      +I  ++ E+    + ADF VLSSGTT  AKL WR
Sbjct: 61  FLRGLKEGSCPIILHPNISGTFQQQIKHVDGELL---KKADFAVLSSGTTGKAKLFWRRL 117
```

```
Query: 116 TSWSDFFSIQNAYFSVTSNSKLFIQGDFSFTGNLNLALSLLLLGGTLVVTQKNSVKYWQT 175
           ++W+  F  QN  F +T NS LF+ G FSFTGNLNLAL+ L  GG LV++QK S+K W +
Sbjct: 118 STWTRLFDYQNKVFGMTGNSCLFLHGSFSFTGNLNLALAQLWAGGCLVLSQKLSLKTWLS 177

Query: 176 LWEKTGVTHLYLLPSYLKLVEQYSKETALDNKTIITSSQYVSDSLLEGLYRKHPKVSVKI 235
           LW+   V+HLYLLP+YL  +  Y  +  +    ++TSSQ +S  LL   Y+K P++ + I
Sbjct: 178 LWQAKKVSHLYLLPTYLNRLLPYLTKNNMTATHLLTSSQMISQELLRHYYKKFPQLEIVI 237

Query: 236 FYGASELNYVSWYDGRDIRDKPQYVGEIVPNVAVRIKEGRIFVKTPYSICGLSSEYCAGD 295
           FYGASEL++++W +GR        VG+  P+V++  K+  IFV+TPYS+ G+S  Y  D
Sbjct: 238 FYGASELSFITWCNGRAAVKINGLVGQPFPDVSISFKDKEIFVETPYSVEGMSQPYSVSD 297

Query: 296 YGELIDGKLYLFGRGGDWCNQSGIKLYLPRLIEKIKTCPYIKDAVAFTKESQSHGQESHC 355
            G++    L L GR  DW NQ G+K +LP L+E     P +K+A A  K  +   +
Sbjct: 298 LGKMSPAGLILEGRQDDWVNQRGVKCHLPSLVELAHQAPNVKEAHAL-KIGKGENETLIL 356

Query: 356 CIVLIENQMQQECLKWLSEHFEKKYGFKHYHIVSKIPLMPSGKIDYQQLKRQL 408
            +VL +          +L+ +        K+Y ++  +PL  +GKI+ + L  ++
Sbjct: 357 VLVLTKKDCLAPIKDFLALYLNSGQLPKYYLVIDCLPLKDNGKINREVLLNKI 409
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 60

A DNA sequence (GBSx0059) was identified in *S.agalactiae* <SEQ ID 187> which encodes the amino acid sequence <SEQ ID 188>. This protein is predicted to be endonuclease III (pdg). Analysis of this protein sequence reveals the following:

```
Possible site: 46

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.00   Transmembrane   25 -  41 (  25 -  41)

----- Final Results -----
               bacterial membrane --- Certainty=0.1001(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB05417 GB:AP001512 endonuclease III (DNA repair) [Bacillus halodurans]
 Identities = 95/202 (47%), Positives = 134/202 (66%)

Query: 1   MLSKAKSRYIIREIIKLFPDAKPSLDFTNVFELLVAVMLSAQTTDAAVNKVTPALFERFP 60
           ML+K +++  +  I  ++PDA+  L  +N FELL+AV+LSAQ TDA VNKVTP LF ++
Sbjct: 1   MLTKKQTQEALAVIADMYPDAECELTHSNPFELLIAVVLSAQCTDALVNKVTPRLFAKYK 60

Query: 61  NPLVLAQADPKEIEPYISKIGLYRNKARFLNQCAKQLIEHFDGKVPRTRQELESLAGVGR 120
            P        +E+E  I  IGLYRNKA+ + +  + L+E + G+VP+ R EL  LAGVGR
Sbjct: 61  TPEDYIAVPLEELEQDIRSIGLYRNKAKNIKKLCQSLLEQYGGEVPQDRDELVKLAGVGR 120

Query: 121 KTANVVMSVGFGIPAFAVDTHVTRICKHHQICKQSASPLEIEKRVMEVLPPEEWLAAHQS 180
           KTANVV SV FG+PA AVDTHV R+ K   IC+   +  ++E+ +M+ +P +EW  +H
Sbjct: 121 KTANVVASVAFGVPAIAVDTHVERVSKRLGICRWKDNVTQVEQTLMKKIPMDEWSISHHR 180

Query: 181 MIYFGRAICHPKNPKCDQYPQL 202
           +I+FGR  C  +NP+CD  P L
Sbjct: 181 LIFFGRYHCKAQNPQCDICPLL 202
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 189> which encodes the amino acid sequence <SEQ ID 190>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
```

```
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 91/199 (45%), Positives = 133/199 (66%)

Query: 2   LSKAKSRYIIREIIKLFPDAKPSLDFTNVFELLVAVMLSAQTTDAAVNKVTPALFERFPN 61
           + KA+   ++  I ++FP+AK  LD+   F+LL+AV+LSAQTTD AVNKVTP L++ +P
Sbjct: 3   IGKARLAKVLTIIGQMFPEAKGELDWETPFQLLIAVILSAQTTDKAVNKVTPGLWQSYPE 62

Query: 62  PLVLAQADPKEIEPYISKIGLYRNKARFLNQCAKQLIEHFDGKVPRTRQELESLAGVGRK 121
              LA A+  ++E  +  IGLY+NKA+ + + A+ + + F G+VP+T +ELESL GVGRK
Sbjct: 63  IEDLAFAELSDVENALRTIGLYKNKAKNIIKTAQAIRDDFKGQVPKTHKELESLPGVGRK 122

Query: 122 TANVVMSVGFGIPAFAVDTHVTRICKHHQICKQSASPLEIEKRVMEVLPPEEWLAAHQSM 181
           TANVV++  +G+PA AVDTHV R+ K   I    A   +IE  +M  +P ++W+  H  +
Sbjct: 123 TANVVLAEVYGVPAIAVDTHVARVSKRLNISSPDADVKQIEADLMAKIPKKDWIITHHRL 182

Query: 182 IYFGRAICHPKNPKCDQYP 200
           I+FGR  C  K PKC+  P
Sbjct: 183 IFFGRYHCLAKKPKCEICP 201
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 61

A DNA sequence (GBSx0060) was identified in *S.agalactiae* <SEQ ID 191> which encodes the amino acid sequence <SEQ ID 192>. Analysis of this protein sequence reveals the following:

```
Possible site: 51

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2264(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA96473 GB:AB036428 hypothetical 8.3 kDa protein [Streptococcus mutans]
 Identities = 53/67 (79%), Positives = 62/67 (92%)

Query: 1  MKVLFDVQNLLKKFGIYVYIGKRLYDIEVMKIELQRLYDNGLISRDDYLKAELILRREHR 60
          MK L+DVQ LLK+FGI+VY+GKRLYDIE+MKIEL+RLYDNGLIS+ DYL AELILRREHR
Sbjct: 1  MKTLYDVQRLLKQFGIFVYLGKRLYDIEMMKIELERLYDNGLISKSDYLHAELILRREHR 60

Query: 61 LELEKEN 67
          +E E+EN
Sbjct: 61 IEKEREN 67
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 193> which encodes the amino acid sequence <SEQ ID 194>. Analysis of this protein sequence reveals the following:

```
Possible site: 57

>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1962(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 53/66 (80%), Positives = 60/66 (90%)

Query: 1  MKVLFDVQNLLKKFGIYVYIGKRLYDIEVMKIELQRLYDNGLISRDDYLKAELILRREHR 60
          MK L+DVQ LLK FGI+VY+GKRLYDIE+MKIELQRLYD+GL+ + DYL AELILRREHR
Sbjct: 7  MKTLYDVQQLLKNFGIFVYLGKRLYDIEMMKIELQRLYDSGLLDKRDYLNAELILRREHR 66

Query: 61 LELEKE 66
          LELEKE
Sbjct: 67 LELEKE 72
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 62

A DNA sequence (GBSx0061) was identified in *S.agalactiae* <SEQ ID 195> which encodes the amino acid sequence <SEQ ID 196>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.06   Transmembrane  133 - 149 ( 133 - 150)

----- Final Results -----
              bacterial membrane --- Certainty=0.1022(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB05144 GB:AP001512 glucose kinase [Bacillus halodurans]
 Identities = 145/315 (46%), Positives = 209/315 (66%), Gaps = 2/315 (0%)

Query: 6   LGIDLGGTTIKFGILTLEGEVQEKWAIETNTLENGRHIVSDIVESLKHRLSLYGLTKDDF 65
           +G+D+GGTTIK   LT  GE+ +KW I TN  + G  I ++I ++L  RLS +  +K D
Sbjct: 7   VGVDVGGTTIKMAFLTTAGEIVDKWEIPTNKQDGGALITTNIADALDKRLSGHHKSKSDL 66

Query: 66  LGIGMGSPGAVDRTSKTVTGAFNLNWADTQEVGSVIEKEVGIPFFIDNDANVAALGERWV 125
           +GIG+G+PG ++  +  +  A N+ W D   +   +E+E  +P  +DNDAN+AALGE W
Sbjct: 67  IGIGLGAPGFIEMDTGFIYHAVNIGWRDFP-LKDKLEEETKLPVIVDNDANIAALGEMWK 125

Query: 126 GAGANNPDVVFVTLGTGVGGGVIADGNLIHGVAGAGGEIGHMIVDPENGFTCTCGNKGCL 185
           GAG    +++ +TLGTGVGGG++A+GN++HGV G  GEIGH+ V PE G  C CG  GCL
Sbjct: 126 GAGDGAKNMLLITLGTGVGGGIVANGNILHGVNGMAGEIGHITVIPEGGAPCNCGKTGCL 185

Query: 186 ETVASATGVVRVARQLAEQYEGSSAIKAAIDNGDTVTSKDIFIAAEDGDKFANSVVERVS 245
           ETVASATG+ R+A +   +++  S +    D    +T+KD+F AA+  D FA SVV+ ++
Sbjct: 186 ETVASATGIARIATEGVTEHK-ESQLALDYDKHGVLTAKDVFSAADASDAFALSVVDHIA 244

Query: 246 RYLGLAAANISNILNPDSVVIGGGVSAAGEFLRSRVEKYFVTFAFPQVKKSTKIKIAELG 305
            YLG A AN++N LNP+ +VIGGGVS AG+ L   ++++F  +A P+V    + +IA LG
Sbjct: 245 YYLGFAIANLANALNPEKIVIGGGVSKAGDTLLKPIKQHFEAYALPRVADGAEFRIATLG 304

Query: 306 NDAGIIGAASLANQQ 320
           NDAG+IG   L  QQ
Sbjct: 305 NDAGVIGGGWLVKQQ 319
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 197> which encodes the amino acid sequence <SEQ ID 198>. Analysis of this protein sequence reveals the following:

```
Possible site: 23

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1060(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 270/319 (84%), Positives = 292/319 (90%)

Query: 1   MSKKLLGIDLGGTTIKFGILTLEGEVQEKWAIETNTLENGRHIVSDIVESLKHRLSLYGL 60
           MS+KLLGIDLGGTTIKFGILT  GEVQEKWAIETN LE G+HIV DI+ S+KHRL LYGL
Sbjct: 1   MSQKLLGIDLGGTTIKFGILTAAGEVQEKWAIETNILEGGKHIVPDIIASIKHRLDLYGL 60

Query: 61  TKDDFLGIGMGSPGAVDRTSKTVTGAFNLNWADTQEVGSVIEKEVGIPFFIDNDANVAAL 120
           +  DF+GIGMGSPGAVDR + TVTGAFNLNW +TQEVGSV+EKE+GIPF IDNDANVAAL
Sbjct: 61  SSADFVGIGMGSPGAVDRDTNTVTGAFNLNWKETQEVGSVVEKELGIPFAIDNDANVAAL 120

Query: 121 GERWVGAGANNPDVVFVTLGTGVGGGVIADGNLIHGVAGAGGEIGHMIVDPENGFTCTCG 180
           GERWVGAG NNPDVVF+TLGTGVGGG+IADGNLIHGVAGAGGEIGHMIV+PENGF CTCG
Sbjct: 121 GERWVGAGENNPDVVFMTLGTGVGGGIIADGNLIHGVAGAGGEIGHMIVEPENGFACTCG 180

Query: 181 NKGCLETVASATGVVRVARQLAEQYEGSSAIKAAIDNGDTVTSKDIFIAAEDGDKFANSV 240
           + GCLETVASATGVV+VAR LAE YEG SAIKAAIDNG+ VTSKDIF+AAE GD FA+SV
Sbjct: 181 SHGCLETVASATGVVKVARLLAEAYEGDSAIKAAIDNGEGVTSKDIFMAAEAGDSFADSV 240

Query: 241 VERVSRYLGLAAANISNILNPDSVVIGGGVSAAGEFLRSRVEKYFVTFAFPQVKKSTKIK 300
           VE+V  YLGLA+ANISNILNPDSVVIGGGVSAAGEFLRSR+EKYFVTF FPQV+ STKIK
Sbjct: 241 VEKVGYYLGLASANISNILNPDSVVIGGGVSAAGEFLRSRIEKYFVTFTFPQVRYSTKIK 300

Query: 301 IAELGNDAGIIGAASLANQ 319
           IAELGNDAGIIGAASLA Q
Sbjct: 301 IAELGNDAGIIGAASLARQ 319
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 63

A DNA sequence (GBSx0062) was identified in *S.agalactiae* <SEQ ID 199> which encodes the amino acid sequence <SEQ ID 200>. Analysis of this protein sequence reveals the following:

```
Possible site: 19

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB14385 GB:Z99116 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 51/124 (41%), Positives = 71/124 (57%), Gaps = 1/124 (0%)

Query: 3   MSVILIIVILLAFVAWASWNYWRVRRAAKFLDNESFQKEMSRGQLIDIREAGAFHRKHIL 62
           MS +++++I  AF+ +   +Y   +R  K L  E F+    + QLID+RE   F   HIL
Sbjct: 1   MSNMIVLIIFPAFIIYMIASYVYQQRIMKTLTEEEFRAGYRKAQLIDVREPNEFEGGHIL 60
```

```
Query: 63  GARNIPASQFKVALSALRKDKPVLLYDASRGQSIPRIVLLLRKEGFNQLYVLKDGFNYWT 122
           GARNIP SQ K   + +R DKPV LY  +  +S  R    LRK G  ++Y LK GF  W
Sbjct: 61  GARNIPLSQLKQRKNEIRTDKPVYLYCQNSVRS-GRAAQTLRKNGCTEIYNLKGGFKKWG 119

Query: 123 GRVK 126
           G++K
Sbjct: 120 GKIK 123
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 201> which encodes the amino acid sequence <SEQ ID 202>. Analysis of this protein sequence reveals the following:

```
     Possible site: 30
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -4.41   Transmembrane    4 -  20 (   1 -  22)

----- Final Results -----
               bacterial membrane --- Certainty=0.2763(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB06532 GB:AP001516 unknown conserved protein [Bacillus halodurans]
 Identities = 46/120 (38%), Positives = 64/120 (53%)

Query: 8   LWLLLVGIVGYYTWNYFSFRKMAKQVDNETFKDVMRQGQLIDLREPAAFRTKHILGARNF 67
           +WL+L+ ++ Y  +      K  K +  E F    R+ QLID+REP  + + HILGARN
Sbjct: 5   VWLVLLALLVYVLFKRLYTPKYLKTLTQEEFIQGYRKAQLIDVREPREYDSGHILGARNI 64

Query: 68  PAQQFDAAIKGLRKDKPVLIYENMRPQYRVPAVKKLKKAGFEDVYVLKDGIDYWDGKVKQ 127
           P  Q    +K +R D+PV +Y     + R  A    KK G EDV  LK G   W GK+K+
Sbjct: 65  PLSQLKQRLKEVRTDQPVYLYCQSGARSRQAAAILKKKHGVEDVNHLKGGFRKWTGKIKK 124
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 63/126 (50%), Positives = 85/126 (67%)

Query: 1   MDMSVILIIVILLAFVAWASWNYWRVRRAAKFLDNESFQKEMSRGQLIDIREAGAFHRKH 60
           M    +++ ++L+  V + +WNY+  R+ AK +DNE+F+  M +GQLID+RE  AF  KH
Sbjct: 1   MSPITLILWLLLVGIVGYYTWNYFSFRKMAKQVDNETFKDVMRQGQLIDLREPAAFRTKH 60

Query: 61  ILGARNIPASQFKVALSALRKDKPVLLYDASRGQSIPRIVLLLRKEGFNQLYVLKDGFNY 120
           ILGARN PA QF  A+  LRKDKPVL+Y+  R Q     V  L+K GF  +YVLKDG +Y
Sbjct: 61  ILGARNFPAQQFDAAIKGLRKDKPVLIYENMRPQYRVPAVKKLKKAGFEDVYVLKDGIDY 120

Query: 121 WTGRVK 126
           W G+VK
Sbjct: 121 WDGKVK 126
```

A related GBS gene <SEQ ID 8483> and protein <SEQ ID 8484> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 1
McG: Discrim Score:     17.55
GvH: Signal Score (-7.5): 3.36
     Possible site: 17
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:   8.86 threshold:  0.0
   PERIPHERAL  Likelihood =  8.86      99
 modified ALOM score:  -2.27

*** Reasoning Step: 3

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
```

```
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
40.4/56.5% over 122aa
Bacillus subtilis
  EGAD|45852| hypothetical 14.6 kd protein in gcvt-spoiiiaa intergenic region Insert
characterized
   SP|P54510|YQHL_BACSU HYPOTHETICAL 14.6 KDA PROTEIN IN GCVT-SPOIIIAA INTERGENIC REGION.
Insert characterized
  GP|1303893|dbj|BAA12549.1||D84432 YqhL Insert characterized
  GP|2634888|emb|CAB14385.1||Z99116 similar to hypothetical proteins Insert characterized
  PIR|C69959|C69959 glpE protein homolog yqhL - Insert characterized

ORF00659(307 - 678 of 978)
EGAD|45852|BS2449(1 - 123 of 126) hypothetical 14.6 kd protein in gcvt-spoiiiaa intergenic
region {Bacillus subtilis}SP|P54510|YQHL_
BACSU     HYPOTHETICAL     14.6     KDA     PROTEIN     IN     GCVT-SPOIIIAA     INTERGENIC
REGION.GP|1303893|dbj|BAA12549.1||D84432 YqhL {Bacillus subtilis}GP|
2634888|emb|CAB14385.1||Z99116     similar     to     hypothetical     proteins     {Bacillus
subtilis}PIR|C69959|C69959 glpE protein homolog yqhL - Bac
illus subtilis
%Match = 13.3
%Identity = 40.3  %Similarity = 56.5
Matches = 50  Mismatches = 53  Conservative Sub.s = 20

108       138       168       198       228       258       288       318
NISNILNPDSVVIGWRCLSSR*IFT*SR*EILCHICFPTS*KVN*N*DC*TR**CWYYWCSKLSQSTSKLRR*GMDMSVI
                                                                          || :
                                                                          MSNM

348       378       408       438       468       498       528       558
LIIVILLAFVAWASWNYWRVRRAAKFLDNESFQKEMSRGQLIDIREAGAFHRKHILGARNIPASQFKVALSALRKDKPVL
::::|: ||: :   :|   :|  | |  | |:    : ||||:||   |   ||||||||| ||:|   : :| ||||
IVLIIFPAFIIYMIASYVYQQRIMKTLTEEEFRAGYRKAQLIDVREPNEFEGGHILGARNIPLSQLKQRKNEIRTDKPVY
               20        30        40        50        60        70        80

588       618       648       678       708       738       768       798
LYDASRGQSIPRIVLLLRKEGFNQLYVLKDGFNYWTGRVK*YTKERVTINNSLHFL*K*IKLKKVENKWHK**NDEKFSY
||          |    ||| |  ::| || ||  | |::|
LY-CQNSVRSGRAAQTLRKNGCTEIYNLKGGFKKWGGKIKAKK
               100       110       120
```

SEQ ID 8484 (GBS13) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 3 (lane 4; MW 16kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 9 (lane 2; MW 40.5kDa).

The GST-fusion protein was purified as shown in Figure 190, lane 5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 64

A DNA sequence (GBSx0063) was identified in *S.agalactiae* <SEQ ID 203> which encodes the amino acid sequence <SEQ ID 204>. This protein is predicted to be regulatory protein TypA (typA). Analysis of this protein sequence reveals the following:

```
Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1738(Affirmative) < succ>
```

```
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB13350 GB:Z99111 similar to GTP-binding elongation factor
          [Bacillus subtilis]
 Identities = 455/609 (74%), Positives = 534/609 (86%), Gaps = 2/609 (0%)

Query: 4   LRTDIRNVAIIAHVDHGKTTLVDELLKQSHTLDERKELEERAMDSNDIEKERGITILAKN 63
           LR D+RN+AIIAHVDHGKTTLVD+LL Q+ T    +++ ERAMDSND+E+ERGITILAKN
Sbjct: 3   LRNDLRNIAIIAHVDHGKTTLVDQLLHQAGTFRANEQVAERAMDSNDLERERGITILAKN 62

Query: 64  TAVAYNDVRINIMDTPGHADFGGEVERIMKMVDGVVLVVDAYEGTMPQTRFVLKKALEQN 123
           TA+ Y D RINI+DTPGHADFGGEVERIMKMVDGVVLVVDAYEG MPQTRFVLKKALEQN
Sbjct: 63  TAINYKDTRINILDTPGHADFGGEVERIMKMVDGVVLVVDAYEGCMPQTRFVLKKALEQN 122

Query: 124 LIPIVVVNKIDKPSARPSEVVDEVLELFIELGADDDQLDFPVVYASAINGTSSMSDDPSD 183
           L P+VVVNKID+  ARP EV+DEVL+LFIEL A+++QL+FPVVYASAINGT+S+  DP
Sbjct: 123 LNPVVVVNKIDRDFARPEEVIDEVLDLFIELDANEEQLEFPVVYASAINGTASL--DPKQ 180

Query: 184 QEKTMAPIFDTIIDHIPAPVDNSEEPLQFQVSLLDYNDFVGRIGIGRVFRGTVKVGDQVT 243
           Q++ M  +++TII H+PAPVDN+EEPLQFQV+LLDYND+VGRIGIGRVFRGT+KVG QV+
Sbjct: 181 QDENMEALYETIIKHVPAPVDNAEEPLQFQVALLDYNDYVGRIGIGRVFRGTMKVGQQVS 240

Query: 244 LSKLDGTTKNFRVTKLFGFFGLERKEIQEAKAGDLIAVSGMEDIFVGETVTPTDAIEPLP 303
           L KLDGT K+FRVTK+FGF GL+R EI+EAKAGDL+AVSGMEDI VGETV P D  +PLP
Sbjct: 241 LMKLDGTAKSFRVTKIFGFQGLKRVEIEEAKAGDLVAVSGMEDINVGETVCPVDHQDPLP 300

Query: 304 VLRIDEPTLQMTFLVNNSPFAGREGKWITSRKVEERLLAELQTDVSLRVDPTDSPDKWTV 363
           VLRIDEPTLQMTF+VNNSPFAGREGK++T+RK+EERL ++LQTDVSLRV+PT SPD W V
Sbjct: 301 VLRIDEPTLQMTFVVNNSPFAGREGKYVTARKIEERLQSQLQTDVSLRVEPTASPDAWVV 360

Query: 364 SGRGELHLSILIETMRREGYELQVSRPEVIIKEIDGVQCEPFERVQIDTPEEYQGAIIQS 423
           SGRGELHLSILIE MRREGYELQVS+PEVIIKEIDGV+CEP ERVQID PEE+ G++++S
Sbjct: 361 SGRGELHLSILIENMRREGYELQVSKPEVIIKEIDGVRCEPVERVQIDVPEEHTGSVMES 420

Query: 424 LSERKGDMLDMQMVGNGQTRLIFLIPARGLIGYSTEFLSMTRGYGIMNHTFDQYLPVVQG 483
           +  RKG+M+DM   GNGQ RLIF +P+RGLIGYSTEFLS+TRG+GI+NHTFD Y P+  G
Sbjct: 421 MGARKGEMVDMINNGNGQVRLIFTVPSRGLIGYSTEFLSLTRGFGILNHTFDSYQPMQAG 480

Query: 484 EIGGRHRGALVSIENGKATTYSIMRIEERGTIFVNPGIEVYEGMIVGENSRDNDLGVNIT 543
           ++GGR +G LVS+ENGKAT+Y I  IE+RG IFV PG EVYEGMIVGE++RDNDL VN++
Sbjct: 481 QVGGRRQGVLVSMENGKATSYGIQGIEDRGVIFVEPGTEVYEGMIVGEHNRDNDLVVNVS 540

Query: 544 TAKQMTNVRSATKDQTAVIKTPRILTLEESLEFLADDEYMEVTPESIRLRKQILNKAARD 603
             KQ TNVRSATKDQT  IK  RI++LEESLE+L +DEY EVTPESIRLRK+ILNK  R+
Sbjct: 541 KMKQQTNVRSATKDQTTTIKKARIMSLEESLEYLNEDEYCEVTPESIRLRKKILNKNERE 600

Query: 604 KANKKKKSA 612
           KA KKKK+A
Sbjct: 601 KAAKKKKTA 609
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 205> which encodes the amino acid sequence <SEQ ID 206>. Analysis of this protein sequence reveals the following:

```
Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1738(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 594/613 (96%), Positives = 607/613 (98%)

Query: 1   MTNLRTDIRNVAIIAHVDHGKTTLVDELLKQSHTLDERKELEERAMDSNDIEKERGITIL 60
           MTNLR DIRNVAIIAHVDHGKTTLVDELLKQSHTLDERKEL+ERAMDSND+EKERGITIL
Sbjct: 1   MTNLRNDIRNVAIIAHVDHGKTTLVDELLKQSHTLDERKELQERAMDSNDLEKERGITIL 60

Query: 61  AKNTAVAYNDVRINIMDTPGHADFGGEVERIMKMVDGVVLVVDAYEGTMPQTRFVLKKAL 120
           AKNTAVAYNDVRINIMDTPGHADFGGEVERIMKMVDGVVLVVDAYEGTMPQTRFVLKKAL
Sbjct: 61  AKNTAVAYNDVRINIMDTPGHADFGGEVERIMKMVDGVVLVVDAYEGTMPQTRFVLKKAL 120

Query: 121 EQNLIPIVVVNKIDKPSARPSEVVDEVLELFIELGADDDQLDFPVVYASAINGTSSMSDD 180
           EQNLIPIVVVNKIDKPSARP+EVVDEVLELFIELGADD+QL+FPVVYASAINGTSS+SDD
Sbjct: 121 EQNLIPIVVVNKIDKPSARPAEVVDEVLELFIELGADDEQLEFPVVYASAINGTSSLSDD 180

Query: 181 PSDQEKTMAPIFDTIIDHIPAPVDNSEEPLQFQVSLLDYNDFVGRIGIGRVFRGTVKVGD 240
           P+DQE TMAPIFDTIIDHIPAPVDNS+EPLQFQVSLLDYNDFVGRIGIGRVFRGTVKVGD
Sbjct: 181 PADQEHTMAPIFDTIIDHIPAPVDNSDEPLQFQVSLLDYNDFVGRIGIGRVFRGTVKVGD 240

Query: 241 QVTLSKLDGTTKNFRVTKLFGFFGLERKEIQEAKAGDLIAVSGMEDIFVGETVTPTDAIE 300
           QVTLSKLDGTTKNFRVTKLFGFFGLER+EIQEAKAGDLIAVSGMEDIFVGET+TPTD +E
Sbjct: 241 QVTLSKLDGTTKNFRVTKLFGFFGLERREIQEAKAGDLIAVSGMEDIFVGETITPTDCVE 300

Query: 301 PLPVLRIDEPTLQMTFLVNNSPFAGREGKWITSRKVEERLLAELQTDVSLRVDPTDSPDK 360
            LP+LRIDEPTLQMTFLVNNSPFAGREGKWITSRKVEERLLAELQTDVSLRVDPTDSPDK
Sbjct: 301 ALPILRIDEPTLQMTFLVNNSPFAGREGKWITSRKVEERLLAELQTDVSLRVDPTDSPDK 360

Query: 361 WTVSGRGELHLSILIETMRREGYELQVSRPEVIIKEIDGVQCEPFERVQIDTPEEYQGAI 420
           WTVSGRGELHLSILIETMRREGYELQVSRPEVIIKEIDGV+CEPFERVQIDTPEEYQGAI
Sbjct: 361 WTVSGRGELHLSILIETMRREGYELQVSRPEVIIKEIDGVKCEPFERVQIDTPEEYQGAI 420

Query: 421 IQSLSERKGDMLDMQMVGNGQTRLIFLIPARGLIGYSTEFLSMTRGYGIMNHTFDQYLPV 480
           IQSLSERKGDMLDMQMVGNGQTRLIFLIPARGLIGYSTEFLSMTRGYGIMNHTFDQYLPV
Sbjct: 421 IQSLSERKGDMLDMQMVGNGQTRLIFLIPARGLIGYSTEFLSMTRGYGIMNHTFDQYLPV 480

Query: 481 VQGEIGGRHRGALVSIENGKATTYSIMRIEERGTIFVNPGIEVYEGMIVGENSRDNDLGV 540
           VQGEIGGRHRGALVSIENGKATTYSIMRIEERGTIFVNPG EVYEGMIVGENSRDNDLGV
Sbjct: 481 VQGEIGGRHRGALVSIENGKATTYSIMRIEERGTIFVNPGTEVYEGMIVGENSRDNDLGV 540

Query: 541 NITTAKQMTNVRSATKDQTAVIKTPRILTLEESLEFLADDEYMEVTPESIRLRKQILNKA 600
           NITTAKQMTNVRSATKDQTAVIKTPRILTLEESLEFL DDEYMEVTPESIRLRKQILNKA
Sbjct: 541 NITTAKQMTNVRSATKDQTAVIKTPRILTLEESLEFLNDDEYMEVTPESIRLRKQILNKA 600

Query: 601 ARDKANKKKKSAE 613
           ARDKANKKKKSAE
Sbjct: 601 ARDKANKKKKSAE 613
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 65

A DNA sequence (GBSx0065) was identified in *S.agalactiae* <SEQ ID 207> which encodes the amino acid sequence <SEQ ID 208>. This protein is predicted to be D-glutamic acid adding enzyme MurD (murD). Analysis of this protein sequence reveals the following:

```
RGD motif 441-443

Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9615> which encodes amino acid sequence <SEQ ID 9616> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC95449 GB:AF068902 D-glutamic acid enzyme MurD [Streptococcus pneumoniae]
 Identities = 341/449 (75%), Positives = 394/449 (86%)

Query: 5   MKTITTFENKKVLVLGLARSGEAAARLLAKLGAIVTVNDGKPFDENPTAQSLLEEGIKVV 64
            MK I  F+NKKVLVLGLA+SGE+AARLL KLGAIVTVNDGKPF++NP AQ LLEEGIKV+
Sbjct: 1   MKVIDQFKNKKVLVLGLAKSGESAARLLDKLGAIVTVNDGKPFEDNPAAQCLLEEGIKVI 60

Query: 65  CGSHPLELLDEDFCYMIKNPGIPYNNPMVKKALEKQIPVLTEVELAYLVSESQLIGITGS 124
             G HPLELLDE+F  M+KNPGIPY+NPM++KAL K IPVLTEVELAYL+SE+ +IGITGS
Sbjct: 61  TGGHPLELLDEEFALMVKNPGIPYSNPMIEKALAKGIPVLTEVELAYLISEAPIIGITGS 120

Query: 125 NGKTTTTTMIAEVLNAGGQRGLLAGNIGFPASEVVQAANDKDTLVMELSSFQLMGVKEFR 184
            NGKTTTTTMI EVL A GQ GLL+GNIG+PAS+V Q A DK+TLVMELSSFQLMGV+EF
Sbjct: 121 NGKTTTTTMIGEVLTAAGQHGLLSGNIGYPASQVAQIATDKNTLVMELSSFQLMGVQEFH 180

Query: 185 PHIAVITNLMPTHLDYHGSFEDYVAAKWNIQNQMSSSDFLVLNFNQGISKELAKTTKATI 244
            P IAVITNLMPTH+DYHG FE+YVAAKWNIQN+M+++DFLVLNFNQ + K+LA  T+AT+
Sbjct: 181 PEIAVITNLMPTHIDYHGLFEEYVAAKWNIQNKMTAADFLVLNFNQDLVKDLASKTEATV 240

Query: 245 VPFSTTEKVDGAYVQDKQLFYKGENIMSVDDIGVPGSHNVENALATIAVAKLAGISNQVI 304
            VPFST EKVDGAY++D QL+++GE +M+ ++IGVPGSHNVENALATIAVAKL G+ NQ I
Sbjct: 241 VPFSTLEKVDGAYLEDGQLYFRGEVVMAANEIGVPGSHNVENALATIAVAKLRGVDNQTI 300

Query: 305 RETLSNFGGVKHRLQSLGKVHGISFYNDSKSTNILATQKALSGFDNTKVILIAGGLDRGN 364
            +ETLS FGGVKHRLQ +  + G+ FYNDSKSTNILATQKALSGFDN+KV+LIAGGLDRGN
Sbjct: 301 KETLSAFGGVKHRLQFVDDIKGVKFYNDSKSTNILATQKALSGFDNSKVVLIAGGLDRGN 360

Query: 365 EFDELIPDITGLKHMVVLGESASRVKRAAQKAGVTYSDALDVRDAVHKAYEVAQQGDVIL 424
            EFDEL+PDITGLK MV+LG+SA RVKRAA KAGV Y +A D+ DA  KAYE+A QGDV+L
Sbjct: 361 EFDELVPDITGLKKMVILGQSAERVKRAADKAGVAYVEATDIADATRKAYELATQGDVVL 420

Query: 425 LSPANASWDMYKNFEVRGDEFIDTFESLR 453
            LSPANASWDMY NFEVRGD FIDT   L+
Sbjct: 421 LSPANASWDMYANFEVRGDLFIDTVAELK 449
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 209> which encodes the amino acid sequence <SEQ ID 210>. Analysis of this protein sequence reveals the following:

```
     Possible site: 25
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 436-438
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 329/451 (72%), Positives = 397/451 (87%)

Query: 5   MKTITTFENKKVLVLGLARSGEAAARLLAKLGAIVTVNDGKPFDENPTAQSLLEEGIKVV 64
            MK I+ F+NKK+L+LGLA+SGEAAA+LL KLGA+VTVND KPFD+NP AQ+LLEEGIKV+
Sbjct: 1   MKVISNFQNKKILILGLAKSGEAAAKLLTKLGALVTVNDSKPFDQNPAAQALLEEGIKVI 60

Query: 65  CGSHPLELLDEDFCYMIKNPGIPYNNPMVKKALEKQIPVLTEVELAYLVSESQLIGITGS 124
            CGSHP+ELLDE+F YM+KNPGIPY+NPMVK+AL K+IP+LTEVELAY VSE+ +IGITGS
Sbjct: 61  CGSHPVELLDENFEYMVKNPGIPYDNPMVKRALAKEIPILTEVELAYFVSEAPIIGITGS 120

Query: 125 NGKTTTTTMIAEVLNAGGQRGLLAGNIGFPASEVVQAANDKDTLVMELSSFQLMGVKEFR 184
```

```
            NGKTTTTTMIA+VLNAGGQ  LL+GNIG+PAS+VVQ A   DTLVMELSSFQL+GV  FR
Sbjct: 121 NGKTTTTTMIADVLNAGGQSALLSGNIGYPASKVVQKAIAGDTLVMELSSFQLVGVNAFR 180

Query: 185 PHIAVITNLMPTHLDYHGSFEDYVAAKWNIQNQMSSSDFLVLNFNQGISKELAKTTKATI 244
            PHIAVITNLMPTHLDYHGSFEDYVAAKW IQ QM+ SD+L+LN NQ IS  LAKTTKAT+
Sbjct: 181 PHIAVITNLMPTHLDYHGSFEDYVAAKWMIQAQMTESDYLILNANQEISATLAKTTKATV 240

Query: 245 VPFSTTEKVDGAYVQDKQLFYKGENIMSVDDIGVPGSHNVENALATIAVAKLAGISNQVI 304
            +PFST + VDGAY++D  L++K + I++  D+GVPGSHN+ENALATIAVAKL+GI++ +I
Sbjct: 241 IPFSTQKVVDGAYLKDGILYFKEQAIIAATDLGVPGSHNIENALATIAVAKLSGIADDII 300

Query: 305 RETLSNFGGVKHRLQSLGKVHGISFYNDSKSTNILATQKALSGFDNTKVILIAGGLDRGN 364
             + LS+FGGVKHRLQ +G++  I+FYNDSKSTNILATQKALSGFDN+++ILIAGGLDRGN
Sbjct: 301 AQCLSHFGGVKHRLQRVGQIKDITFYNDSKSTNILATQKALSGFDNSRLILIAGGLDRGN 360

Query: 365 EFDELIPDITGLKHMVVLGESASRVKRAAQKAGVTYSDALDVRDAVHKAYEVAQQGDVIL 424
            EFD+L+PD+ GLK M++LGESA R+KRAA KA V+Y +A +V +A   A+++AQ GD IL
Sbjct: 361 EFDDLVPDLLGLKQMIILGESAERMKRAANKAEVSYLEARNVAEATELAFKLAQTGDTIL 420

Query: 425 LSPANASWDMYKNFEVRGDEFIDTFESLRGE 455
            LSPANASWDMY NFEVRGDEF+ TF+ LRG+
Sbjct: 421 LSPANASWDMYPNFEVRGDEFLATFDCLRGD 451
```

SEQ ID 208 (GBS305) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 11; MW 53.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 56 (lane 3; MW 79kDa).

The GBS305-GST fusion product was purified (Figure 207, lane 8) and used to immunise mice. The resulting antiserum was used for FACS (Figure 270), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 66

A DNA sequence (GBSx0066) was identified in *S.agalactiae* <SEQ ID 211> which encodes the amino acid sequence <SEQ ID 212>. Analysis of this protein sequence reveals the following:

```
RGD motif 285-287

Possible site: 60

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.65   Transmembrane   74 -  90 (  73 -  93)

----- Final Results -----
               bacterial membrane --- Certainty=0.1659(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 213> which encodes the amino acid sequence <SEQ ID 214>. Analysis of this protein sequence reveals the following:

```
Possible site: 37

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.33   Transmembrane   81 -  97 (  80 - 100)
   INTEGRAL    Likelihood = -0.16   Transmembrane  272 - 288 ( 271 - 288)

----- Final Results -----
               bacterial membrane --- Certainty=0.1532(Affirmative) < succ>
```

```
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9141> which encodes the amino acid sequence <SEQ ID 9142>. Analysis of this protein sequence reveals the following:

```
      Possible site: 37

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.33   Transmembrane   74 -  90
   INTEGRAL    Likelihood = -0.16   Transmembrane  265 - 281

----- Final Results -----
              bacterial membrane --- Certainty=0.1532(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 286-288
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 249/358 (69%), Positives = 293/358 (81%), Gaps = 1/358 (0%)

Query: 1   MGKKIVFTGGGTVGHVTLNLILIPKFIKDGWEVHYIGDKNGIEHEQINQSGLDITFHSIA 60
           M KKI+FTGGGTVGHVTLNLILIPKFIKDGWEVHYIGDKNGIEH +I +SGLD+TFH+IA
Sbjct: 8   MPKKILFTGGGTVGHVTLNLILIPKFIKDGWEVHYIGDKNGIEHTEIEKSGLDVTFHAIA 67

Query: 61  TGKLRRYFSWQNMLDVFKVGVGVLQSIAIIAKLRPQALFSKGGFVSVPPVVAARLLKVPV 120
           TGKLRRYFSWQN+ DVFKV +G+LQS+ I+AKLRPQALFSKGGFVSVPPVVAA+LL  PV
Sbjct: 68  TGKLRRYFSWQNLADVFKVALGLLQSLFIVAKLRPQALFSKGGFVSVPPVVAAKLLGKPV 127

Query: 121 FVHESDLSMGLANKIAYKFATIMYTTFEQSKDLIKTKHIGAVTKVM-DCKKSFENTDLTS 179
           F+HESD SMGLANKIAYKFAT MYTTFEQ   L K KH+GAVTKV  D  +  E+T L +
Sbjct: 128 FIHESDRSMGLANKIAYKFATTMYTTFEQEDQLSKVKHLGAVTKVFKDANQMPESTQLEA 187

Query: 180 IKEAFDPNLKTLLFIGGSAGAKVFNDFITQTPELEEKYNVINISGDSSLNRLKKNLYRVD 239
           +KE F  +LKTLLFIGGSAGA VFN FI+  PEL+++YN+INI+GD  LN L  +LYRVD
Sbjct: 188 VKEYFSRDLKTLLFIGGSAGAHVFNQFISDHPELKQRYNIIINITGDPHLNELSSHLYRVD 247

Query: 240 YVTDLYQPLMNLADVVVTRGGSNTIFELVAMKKLHLIIPLGREASRGDQLENAAYFEEKG 299
           YVTDLYQPLM +AD+VVTRGGSNT+FEL+AM KLHLI+PLG+EASRGDQLENA YFE++G
Sbjct: 248 YVTDLYQPLMAMADLVVTRGGSNTLFELLAMAKLHLIVPLGKEASRGDQLENATYFEKRG 307

Query: 300 YALQLPESELNINTLEKQINLLISNSESYEKNMSQSSEIKSQDEFYQLLIDDMAKVTK 357
           YA QL E +L ++  ++ +  L  +   YE  M  + EI+S D FY LL  D++   K
Sbjct: 308 YAKQLQEPDLTLHNFDQAMADLFEHQADYEATMLATKEIQSPDFFYDLLRADISSAIK 365
```

SEQ ID 212 (GBS306) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 12; MW 43kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 56 (lane 4; MW 68kDa).

GBS306-GST was purified as shown in Figure 207, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 67

A DNA sequence (GBSx0067) was identified in *S.agalactiae* <SEQ ID 215> which encodes the amino acid sequence <SEQ ID 216>. This protein is predicted to be cell division protein DivIB. Analysis of this protein sequence reveals the following:

```
Possible site: 58
```

```
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-14.33   Transmembrane  103 - 119 (  96 - 124)

----- Final Results -----
               bacterial membrane --- Certainty=0.6731(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC95451 GB:AF068902 cell division protein DivIB [Streptococcus pneumoniae]
 Identities = 119/396 (30%), Positives = 214/396 (53%), Gaps = 38/396 (9%)

Query: 3   KKKSDTPEKEEVV-LTEWQKRNLEFLKKRKEDEE---EQKRINEKLRLDKRS-----KLN 53
           KK  D    EE+  L+EWQKRN E+LKK+ E+E    E+K    + R+ + S     K +
Sbjct: 5   KKNEDKEILEELKELSEWQKRNQEYLKKKAEEEAALAEEKEKERQARMGEESEKSEDKQD 64

Query: 54  ISSPEEPQNTTKIKKLHFPKIS------------RPKIEKKQKKEKIVNSLAKTNR---- 97
             S  + +++   K+    K++              P+ ++K++++K ++  A   +
Sbjct: 65  QESETDQEDSESAKEESEEKVASSEADKEKEEKEEPESKEKEEQDKKLSKKATKEKPAKA 124

Query: 98  -------IRTAPIFVVAFLVILVSVFLLTPFSKQKTITVSGNQHTPDDILIEKTNIQKND 150
                 +R   I   + L+++VS +LL+P++  K I V G   T  D + + + IQ +D
Sbjct: 125 KIPGIHILRAFTILFPSLLLLIVSAYLLSPYATMKDIRVEGTVQTTADDIRQASGIQDSD 184

Query: 151 YFFSLIFKHKAIEQRLAAEDVWVKTAQMTYQFPNKFHIQVQENKIIAYAHTKQGYQPVLE 210
           Y  +L+      E+++ + + WV++AQ+ YQFP KF I+V+E  I+AY  + + + P+L
Sbjct: 185 YTINLLLDKAKYEKQIKS-NYWVESAQLVYQFPTKFTIKVKEYDIVAYYISGENHYPILS 243

Query: 211 TGK-KADPVNSSELPKHFLTINLDKEDSIKLLIKDLKALDPDLISEIQVISLADSKTTPD 269
           +G+ +   V+ + LP+ +L++  +  + IK+ + +L  + P+L + IQ + LA SK T D
Sbjct: 244 SGQLETSSVSLNSLPETYLSVLFNDSEQIKVFVSELAQISPELKAAIQKVELAPSKVTSD 303

Query: 270 LLLLDMHDGNSIRIPLSKFKERLPFYKQIKKNLKEPSIVDMEVGVYTTTNTIESTPVKAE 329
           L+ L M+D + + +PLS+  ++LP+Y +IK  L EPS+VDME G+Y+ T   +     E
Sbjct: 304 LIRLTMNDSDEVLVPLSEMSKKLPYYSKIKPQLSEPSVVDMEAGIYSYTVADKLIMEVEE 363

Query: 330 DTKNKSTDKTQTQNGQVAENSQGQTNNSNTNQQGQQ 365
             K ++ +  + Q     E  + Q   SN NQ  Q+
Sbjct: 364 KAKQEAKEAEKKQE----EEQKKQEEESNRNQTTQR 395
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 217> which encodes the amino acid sequence <SEQ ID 218>. Analysis of this protein sequence reveals the following:

```
     Possible site: 59
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.45   Transmembrane  106 - 122 ( 102 - 125)

----- Final Results -----
               bacterial membrane --- Certainty=0.4779(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 152/381 (39%), Positives = 232/381 (59%), Gaps = 14/381 (3%)

Query: 4   KKSDTPEKEEVVLTEWQKRNLEFLKKRKEDEEEQKRINEKLRLDKRSKLNISSPEEP--- 60
           K  +    +++VLTEWQKRN+EFLKK+K+  EE+K++ EKL  DK+++    +  E
Sbjct: 3   KDKEKQSDDKLVLTEWQKRNIEFLKKKKQQAEEEKKLKEKLLSDKKAQQQAQNASEAVEL 62

Query: 61  --QNTTKIKKLHFPKISRPKIEKK--QKKEKIVNSLAKTNRIRTAPIFVVAFLVILVSVF 116
               T  +++     S+PK  KK  Q KEK     +A      ++ P+ + A L++ VS+F
Sbjct: 63  KTDEKTDSQEIESETTSKPKKTKKVRQPKEKSATQIAFQ---KSLPVLLGALLLMAVSIF 119

Query: 117 LLTPFSKQKTITVSGNQHTPDDILIEKTNIQKNDYFFSLIFKHKAIEQRLAAEDVWVKTA 176
```

```
            ++TP+SK+K  +V GN  T  D LI+ + ++ +DY+ +L+      E+ +     WVK+
Sbjct: 120 MITPYSKKKEFSVRGNHQTNLDELIKASKVKASDYWLTLLTSPGQYERPILRTIPWVKSV 179

Query: 177 QMTYQFPNKFHIQVQENKIIAYAHTKQGYQPVLETGKKADPVNSSELPKHFLTINLDKED 236
            ++YQFPN F   V E +IIAYA  + G+QP+LE GK+ D V +SELPK FL +NL  E
Sbjct: 180 HLSYQFPNHFLFNVIEFEIIAYAQVENGFQPILENGKRVDKVRASELPKSFLILNLKDEK 239

Query: 237 SIKLLIKDLKALDPDLISEIQVISLADSKTTPDLLLLDMHDGNSIRIPLSKFKERLPFYK 296
            +I+ L+K L  L   L+  I+ +SLA+SKTT DLLL++MHDGN +R+P S+   +LP+Y+
Sbjct: 240 AIQQLVKQLTTLPKKLVKNIKSVSLANSKTTADLLLIEMHDGNVVRVPQSQLTLKLPYYQ 299

Query: 297 QIKKNLKEPSIVDMEVGVYTTTNTIESTPVKAEDTKNKSTDKTQTQNGQVAENSQGQTNN 356
            ++KKNL+  SIVDMEVG+YTTT  IE+ P      +  + DK   + G+     Q QT+N
Sbjct: 300 KLKKNLENDSIVDMEVGIYTTTQEIENQPEVPLTPEQNAADKEGDKPGE----HQEQTDN 355

Query: 357 SNTNQQGQQIATEQAPNPQNV 377
            +     Q    +  P+P+ V
Sbjct: 356 DSETPANQSSPQQTPPSPETV 376
```

SEQ ID 216 (GBS85) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 10; MW 45.2kDa).

The GBS85-His fusion product was purified (Figure 105A; see also Figure 193, lane 5) and used to immunise mice (lane 1 product; 20μg/mouse). The resulting antiserum was used for Western blot (Figure 105B), FACS (Figure 105C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 68

A DNA sequence (GBSx0068) was identified in *S.agalactiae* <SEQ ID 219> which encodes the amino acid sequence <SEQ ID 220>. This protein is predicted to be cell division protein FtsA (ftsA). Analysis of this protein sequence reveals the following:

```
Possible site: 56

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -3.19   Transmembrane  322 - 338 ( 321 - 338)

----- Final Results -----
               bacterial membrane --- Certainty=0.2275(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC95439 GB:AF068901 cell division protein FtsA [Streptococcus pneumoniae]
 Identities = 292/457 (63%), Positives = 366/457 (79%), Gaps = 1/457 (0%)

Query: 1   MARNGFFTGLDIGTSSIKVLVAEFIANEMNVIGVSNVPSSGVKDGIIIDIEAAATAIKEA 60
            MAR GFFTGLDIGTSS+KVLVAE    E+NVIGVSN  S GVKDGII+DI+AAATAIK A
Sbjct: 1   MAREGFFTGLDIGTSSVKVLVAEQRNGELNVIGVSNAKSKGVKDGIIVDIDAAATAIKSA 60

Query: 61  VKQAEEKAGITIDKINVGLPANLLQIEPTQGMIPVPNESKEIKDEDVESVVKSALTKSIT 120
            + QAEEKAGI+I  +NVGLP NLLQ+EPTQGMIPV +++KEI D+DVE+VVKSALTKS+T
Sbjct: 61  ISQAEEKAGISIKSVNVGLPGNLLQVEPTQGMIPVTSDTKEITDQDVENVVKSALTKSMT 120

Query: 121 PEREVISLIPLEFIVDGFQGIRDPRGMMGIRLEMRGLIYTGPTTILHNLRKTVERAGIKV 180
            P+REVI+ IP EFIVDGFQGIRDPRGMMG+RLEMRGL+YTGP TILHNLRKTVERAG++V
Sbjct: 121 PDREVITFIPEEFIVDGFQGIRDPRGMMGVRLEMRGLLYTGPRTILHNLRKTVERAGVQV 180
```

```
Query: 181 EHVVIAPLALAKSVLNEGEREFGATVIDMGGGQTTVASMRNQELQYTNIYSEGSDYVTKD 240
           E+V+I+PLA+ +SVLNEGEREFGATVIDMG GQTTVA++RNQELQ+T+I  EG DYVTKD
Sbjct: 181 ENVIISPLAMVQSVLNEGEREFGATVIDMGAGQTTVATIRNQELQFTHILQEGGDYVTKD 240

Query: 241 ISKVLRTTVEIAEALKFNFGQANVEEASTSDTVQVNVVGNEEPVEITESYLSQIISGRIR 300
           ISKVL+T+ ++AE LK N+G+A    AS  +T QV V+G  E VE+TE+YLS+IIS RI+
Sbjct: 241 ISKVLKTSRKLAEGLKLNYGEAYPPLAS-KETFQVEVIGEVEAVEVTEAYLSEIISARIK 299

Query: 301 QILEHVKQDLGRGRLLDLPGGIILVGGGAIMPGVVEVAQQIFGTRVKLHVPNQVGIRNPM 360
            ILE +KQ+L R RLLDLPGGI+L+GG AI+PG+VE+AQ++FG RVKL+VPNQVGIRNP
Sbjct: 300 HILEQIKQELDRRRLLDLPGGIVLIGGNAILPGMVELAQEVFGVRVKLYVPNQVGIRNPA 359

Query: 361 FANVISIVDYVGMMSEVDIIAQHAVTGDEMLRHKPVDFDYKEKTNTMSTMPYSEPLTSSM 420
           FA+VIS+ ++ G ++EV+++AQ A+ G+  L H+P+ F    +                +
Sbjct: 360 FAHVISLSEFAGQLTEVNLLAQGAIKGENDLSHQPISFGGMLQKTAQFVQSTPVQPAPAP 419

Query: 421 EDSNLEPIRARENAQEPTEPKANIGERIRGIFGSMFD 457
           E   + P     + Q+ ++ K  + +R RG+ GSMFD
Sbjct: 420 EVEPVAPTEPMADFQQASQNKPKLADRFRGLIGSMFD 456
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 221> which encodes the amino acid sequence <SEQ ID 222>. Analysis of this protein sequence reveals the following:

```
     Possible site: 55

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.35   Transmembrane  313 - 329 ( 312 - 329)

----- Final Results -----
               bacterial membrane --- Certainty=0.2338(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC95439 GB:AF068901 cell division protein FtsA [Streptococcus pneumoniae]
 Identities = 299/448 (66%), Positives = 368/448 (81%), Gaps = 4/448 (0%)

Query: 1   LDIGTSSIKVLVAEFISGEMNVIGVSNVPSTGVKDGIIIDIEAAATAIKTAVEQAEEKAG 60
           LDIGTSS+KVLVAE  +GE+NVIGVSN  S GVKDGII+DI+AAATAIK+A+ QAEEKAG
Sbjct: 10  LDIGTSSVKVLVAEQRNGELNVIGVSNAKSKGVKDGIIVDIDAAATAIKSAISQAEEKAG 69

Query: 61  MTIEKVNVGLPANLLQIEPTQGMIPVPSESKEIKDEDVDSVVKSALTKSITPEREVISLV 120
           ++I+ VNVGLP NLLQ+EPTQGMIPV S++KEI D+DV++VVKSALTKS+TP+REVI+ +
Sbjct: 70  ISIKSVNVGLPGNLLQVEPTQGMIPVTSDTKEITDQDVENVVKSALTKSMTPDREVITFI 129

Query: 121 PEEFIVDGFQGIRDPRGMMGIRLEMRGLIYTGPSTILHNLRKTVERAGIKVENIIISPLA 180
           PEEFIVDGFQGIRDPRGMMG+RLEMRGL+YTGP TILHNLRKTVERAG++VEN+IISPLA
Sbjct: 130 PEEFIVDGFQGIRDPRGMMGVRLEMRGLLYTGPRTILHNLRKTVERAGVQVENVIISPLA 189

Query: 181 MAKTILNEGEREFGATVIDMGGGQTTVASMRAQELQYTNIYAEGGEYITKDISKVLKTSL 240
           M +++LNEGEREFGATVIDMG GQTTVA++R QELQ+T+I  EGG+Y+TKDISKVLKTS
Sbjct: 190 MVQSVLNEGEREFGATVIDMGAGQTTVATIRNQELQFTHILQEGGDYVTKDISKVLKTSR 249

Query: 241 AIAEALKFNFGQAEISEASITETVKVDVVGSEEPVEVTERYLSEIISARIRHILDRVKQD 300
            +AE LK N+G+A    AS  ET +V+V+G  E VEVTE YLSEIISARI+HIL+++KQ+
Sbjct: 250 KLAEGLKLNYGEAYPPLAS-KETFQVEVIGEVEAVEVTEAYLSEIISARIKHILEQIKQE 308

Query: 301 LERGRLLDLPGGIVLIGGGAIMPGVVEIAQEIFGVTVKLHVPNQVGIRNPMFSNVISLVE 360
           L+R RLLDLPGGIVLIGG AI+PG+VE+AQE+FGV VKL+VPNQVGIRNP F++VISL E
Sbjct: 309 LDRRRLLDLPGGIVLIGGNAILPGMVELAQEVFGVRVKLYVPNQVGIRNPAFAHVISLSE 368

Query: 361 YVGMMSEVDVLAQTAVSGEELLRRKPIDFSGQESYLPDYDDSRRPESTIGYEQQ---ASQ 417
           + G ++EV++LAQ A+ GE  L  +PI F G       +  S   +     E +    ++
Sbjct: 369 FAGQLTEVNLLAQGAIKGENDLSHQPISFGGMLQKTAQFVQSTPVQPAPAPEVEPVAPTE 428
```

```
Query: 418 TAYDSQVPSDPKQKISERVRGIFGSMFD 445
           D Q  S  K K+++R RG+ GSMFD
Sbjct: 429 PMADFQQASQNKPKLADRFRGLIGSMFD 456
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 349/456 (76%), Positives = 402/456 (87%), Gaps = 19/456 (4%)

Query: 10  LDIGTSSIKVLVAEFIANEMNVIGVSNVPSSGVKDGIIIDIEAAATAIKEAVKQAEEKAG 69
           LDIGTSSIKVLVAEFI+ EMNVIGVSNVPS+GVKDGIIIDIEAAATAIK AV+QAEEKAG
Sbjct: 1   LDIGTSSIKVLVAEFISGEMNVIGVSNVPSTGVKDGIIIDIEAAATAIKTAVEQAEEKAG 60

Query: 70  ITIDKINVGLPANLLQIEPTQGMIPVPNESKEIKDEDVESVVKSALTKSITPEREVISLI 129
           +TI+K+NVGLPANLLQIEPTQGMIPVP+ESKEIKDEDV+SVVKSALTKSITPEREVISL+
Sbjct: 61  MTIEKVNVGLPANLLQIEPTQGMIPVPSESKEIKDEDVDSVVKSALTKSITPEREVISLV 120

Query: 130 PLEFIVDGFQGIRDPRGMMGIRLEMRGLIYTGPTTILHNLRKTVERAGIKVEHVVIAPLA 189
           P EFIVDGFQGIRDPRGMMGIRLEMRGLIYTGP+TILHNLRKTVERAGIKVE+++I+PLA
Sbjct: 121 PEEFIVDGFQGIRDPRGMMGIRLEMRGLIYTGPSTILHNLRKTVERAGIKVENIIISPLA 180

Query: 190 LAKSVLNEGEREFGATVIDMGGGQTTVASMRNQELQYTNIYSEGSDYVTKDISKVLRTTV 249
           +AK++LNEGEREFGATVIDMGGGQTTVASMR QELQYTNIY+EG +Y+TKDISKVL+T++
Sbjct: 181 MAKTILNEGEREFGATVIDMGGGQTTVASMRAQELQYTNIYAEGGEYITKDISKVLKTSL 240

Query: 250 EIAEALKFNFGQANVEEASTSDTVQVNVVGNEEPVEITESYLSQIISGRIRQILEHVKQD 309
            IAEALKFNFGQA + EAS ++TV+V+VVG+EEPVE+TE YLS+IIS RIR IL+ VKQD
Sbjct: 241 AIAEALKFNFGQAEISEASITETVKVDVVGSEEPVEVTERYLSEIISARIRHILDRVKQD 300

Query: 310 LGRGRLLDLPGGIILVGGGAIMPGVVEVAQQIFGTRVKLHVPNQVGIRNPMFANVISIVD 369
           L RGRLLDLPGGI+L+GGGAIMPGVVE+AQ+IFG  VKLHVPNQVGIRNPMF+NVIS+V+
Sbjct: 301 LERGRLLDLPGGIVLIGGGAIMPGVVEIAQEIFGVTVKLHVPNQVGIRNPMFSNVISLVE 360

Query: 370 YVGMMSEVDIIAQHAVTGDEMLRHKPVDF--------DYKEKTNTMSTMPYSEPLTSSME 421
           YVGMMSEVD++AQ AV+G+E+LR KP+DF         DY +     ST+ Y +  + +
Sbjct: 361 YVGMMSEVDVLAQTAVSGEELLRRKPIDFSGQESYLPDYDDSRRPESTIGYEQQASQTAY 420

Query: 422 DSNLEPIRARENAQEPTEPKANIGERIRGIFGSMFD 457
           DS            Q P++PK  I ER+RGIFGSMFD
Sbjct: 421 DS-----------QVPSDPKQKISERVRGIFGSMFD 445
```

SEQ ID 220 (GBS73) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 5; MW 47.8kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 20 (lane 5; MW 70.1kDa).

GBS73-GST was purified as shown in Figure 197, lane 7.

The GBS73-His fusion product was purified (Figure 103A) and used to immunise mice (lane 1 product; 20μg/mouse). The resulting antiserum was used for Western blot (Figure 103B), FACS (Figure 103C) and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 69

A DNA sequence (GBSx0069) was identified in *S.agalactiae* <SEQ ID 223> which encodes the amino acid sequence <SEQ ID 224>. This protein is predicted to be cell division protein FtsZ (ftsz). Analysis of this protein sequence reveals the following:

```
Possible site: 56
```

```
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -1.97   Transmembrane  117 - 133 ( 117 - 133)

----- Final Results -----
               bacterial membrane --- Certainty=0.1786(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC95440 GB:AF068901 cell division protein FtsZ [Streptococcus pneumoniae]
 Identities = 327/426 (76%), Positives = 363/426 (84%), Gaps = 7/426 (1%)

Query: 1   MVFSFDTASVQGAVIKVIGVGGGGGNAINRMIDEGVAGVEFIAANTDIQALSSSKAETVI 60
            M FSFDTA+ QGAVIKVIGVGGGGGNAINRM+DEGV GVEFIAANTD+QALSS+KAETVI
Sbjct: 1   MTFSFDTAAAQGAVIKVIGVGGGGGNAINRMVDEGVTGVEFIAANTDVQALSSTKAETVI 60

Query: 61  QLGPKLTRGLGAGGQPEVGRKAAEESEEVLTEALTGADMVFITAGMGGGSGTGAAPVIAR 120
            QLGPKLTRGLGAGGQPEVGRKAAEESEE LTEA++GADMVFITAGMGGGSGTGAAPVIAR
Sbjct: 61  QLGPKLTRGLGAGGQPEVGRKAAEESEETLTEAISGADMVFITAGMGGGSGTGAAPVIAR 120

Query: 121 IAKSLGALTVAVITRPFGFEGNKRSNFAIEGIQELREQVDTLLIISNNNLLEIVDKKTPL 180
            IAK LGALTV V+TRPFGFEG+KR  FA+EGI +LRE VDTLLIISNNNLLEIVDKKTPL
Sbjct: 121 IAKDLGALTVGVVTRPFGFEGSKRGQFAVEGINQLREHVDTLLIISNNNLLEIVDKKTPL 180

Query: 181 LEALSEADNVLRQGVQGITDLITNPGLINLDFADVKTVMANKGNALMGIGIGSGEERITE 240
            LEALSEADNVLRQGVQGITDLITNPGLINLDFADVKTVMANKGNALMGIGIGSGEER+ E
Sbjct: 181 LEALSEADNVLRQGVQGITDLITNPGLINLDFADVKTVMANKGNALMGIGIGSGEERVVE 240

Query: 241 AARKAIYSPLLETTIDGAEDVIVNVTGGMDMTLTEAEEASEIVSQAAGKGVNIWLGTSID 300
            AARKAIYSPLLETTIDGAEDVIVNVTGG+D+TL EAEEAS+IV+QAAG+GVNIWLGTSID
Sbjct: 241 AARKAIYSPLLETTIDGAEDVIVNVTGGLDLTLIEAEEASQIVNQAAGQGVNIWLGTSID 300

Query: 301 MDMKDEIRVTVVATGVRKDKTNQVSGFTTSAPTNQAPSERQSTSNSNFDRRGNFDMTESR 360
              M+DEIRVTVVATGVR+D+  +V       + TN   + + + S+  FDR  +FDM E+
Sbjct: 301 ESMRDEIRVTVVATGVRQDRVEKVVAPQARSATNYRETVKPAHSH-GFDR--HFDMAETA 357

Query: 361 EMPTQQNQPHAQNQQQSSAFGNWDLRRDNISRPTEGELDSKLSMSTFSENDDMDDELETP 420
            E+P Q   P      Q+SAFG+WDLRR++I R T+  +            D  +DEL+TP
Sbjct: 358 ELPKQ--NPRRLEPTQASAFGDWDLRRESIVRTTDSVVSPVERFEAPISQD--EDELDTP 413

Query: 421 PFFKNR 426
            PFFKNR
Sbjct: 414 PFFKNR 419
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 225> which encodes the amino acid sequence <SEQ ID 226>. Analysis of this protein sequence reveals the following:

```
Possible site: 56

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -1.81   Transmembrane  117 - 133 ( 117 - 133)

----- Final Results -----
               bacterial membrane --- Certainty=0.1723(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 372/439 (84%), Positives = 391/439 (88%), Gaps = 13/439 (2%)

Query: 1   MVFSFDTASVQGAVIKVIGVGGGGGNAINRMIDEGVAGVEFIAANTDIQALSSSKAETVI 60
            M FSFDTAS+QGA+IKVIGVGGGGGNAINRMIDEGVAGVEFIAANTDIQALSSSKAETVI
Sbjct: 1   MAFSFDTASIQGAIIKVIGVGGGGGNAINRMIDEGVAGVEFIAANTDIQALSSSKAETVI 60
```

```
Query: 61  QLGPKLTRGLGAGGQPEVGRKAAEESEEVLTEALTGADMVFITAGMGGGSGTGAAPVIAR 120
           QLGPKLTRGLGAGGQPEVGRKAAEESEE+LTEALTGADMVFITAGMGGGSGTGAAPVIAR
Sbjct: 61  QLGPKLTRGLGAGGQPEVGRKAAEESEEILTEALTGADMVFITAGMGGGSGTGAAPVIAR 120

Query: 121 IAKSLGALTVAVITRPFGFEGNKRSNFAIEGIQELREQVDTLLIISNNNLLEIVDKKTPL 180
           IAKSLGALTVAV+TRPFGFEGNKR NFAIEGI+ELREQVDTLLIISNNNLLEIVDKKTPL
Sbjct: 121 IAKSLGALTVAVVTRPFGFEGNKRGNFAIEGIEELREQVDTLLIISNNNLLEIVDKKTPL 180

Query: 181 LEALSEADNVLRQGVQGITDLITNPGLINLDFADVKTVMANKGNALMGIGIGSGEERITE 240
           LEALSEADNVLRQGVQGITDLIT+PGLINLDFADVKTVMANKGNALMGIGIGSGEERI E
Sbjct: 181 LEALSEADNVLRQGVQGITDLITSPGLINLDFADVKTVMANKGNALMGIGIGSGEERIVE 240

Query: 241 AARKAIYSPLLETTIDGAEDVIVNVTGGMDMTLTEAEEASEIVSQAAGKGVNIWLGTSID 300
           AARKAIYSPLLETTIDGA+DVIVNVTGG+DMTLTEAEEASEIV QAAG+GVNIWLGTSID
Sbjct: 241 AARKAIYSPLLETTIDGAQDVIVNVTGGLDMTLTEAEEASEIVGQAAGQGVNIWLGTSID 300

Query: 301 MDMKDEIRVTVVATGVRKDKTNQVSGF---TTSAPTN--------QAPSERQSTSNSNFD 349
             MKD+IRVTVVATGVR++K  QVSGF    T   TN         A  + + +    FD
Sbjct: 301 DTMKDDIRVTVVATGVRQEKAEQVSGFRQPRTFTQTNAQQVAGAQYASDQAKQSVQPGFD 360

Query: 350 RRGN--FDMTESREMPTQQNQPHAQNQQQSSAFGNWDLRRDNISRPTEGELDSKLSMSTF 407
           RR N  FDM ESRE+P+ Q      NQ Q SAFGNWDLRRDNISRPTEGELD+ L+MSTF
Sbjct: 361 RRSNFDFDMGESREIPSAQKVISNHNQNQGSAFGNWDLRRDNISRPTEGELDNHLNMSTF 420

Query: 408 SENDDMDDELETPPFFKNR 426
           S NDD DDELETPPFFKNR
Sbjct: 421 SANDDSDDELETPPFFKNR 439
```

SEQ ID 224 (GBS163) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 28 (lane 7; MW 44kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 34 (lane 4; MW 69kDa).

The GBS163-GST fusion product was purified (Figure 114A; see also Figure 198, lane 11) and used to immunise mice (lane 1 product; 20μg/mouse). The resulting antiserum was used for Western blot (Figure 114B), FACS and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 70

A DNA sequence (GBSx0070) was identified in *S.agalactiae* <SEQ ID 227> which encodes the amino acid sequence <SEQ ID 228>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2750(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC95441 GB:AF068901 YlmE [Streptococcus pneumoniae]
 Identities = 140/223 (62%), Positives = 177/223 (78%)

Query: 2   MNLQENKTAIFDNVSKLALKAGRAHESVHIVAVTKYVNCQTTEALIRTGVNHIGENRVDK 61
           MN++EN   +F  V++ +L A R   SV ++AVTKYV+  T EAL+  GV+HIGENRVDK
Sbjct: 1   MNVKENTELVFREVAEASLSAHRESGSVSVIAVTKYVDVPTAEALLPLGVHHIGENRVDK 60
```

```
Query: 62  FLEKYQALKDEKLTWHLIGSLQRRKVKDVINYVDYFHALDSVKLAAEIQKHAQKLIKCFL 121
           FLEKY+ALKD  +TWHLIG+LQRRKVKDVI YVDYFHALDSVKLA EIQK + ++IKCFL
Sbjct: 61  FLEKYEALKDRDVTWHLIGTLQRRKVKDVIQYVDYFHALDSVKLAGEIQKRSDRVIKCFL 120

Query: 122 QVNISREDSKHGFTIEQIDDALNLISRYDKIELIGIMTMAPLKATKEEISSIFEETESLR 181
           QVNIS+E+SKHGF+ E++ + L  ++R DKIE +G+MTMAP +A+ E++  IF+  + L+
Sbjct: 121 QVNISKEESKHGFSREELLEILPELARLDKIEYVGLMTMAPFEASSEQLKEIFKAAQDLQ 180

Query: 182 KRLQARNIERMPFTELSMGMSRDYDIAIQNGSTFVRIGTSFFK 224
           + +Q + I  MP TELSMGMSRDY  AIQ GSTFVRIGTSFFK
Sbjct: 181 REIQEKQIPNMPMTELSMGMSRDYKEAIQFGSTFVRIGTSFFK 223
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 229> which encodes the amino acid sequence <SEQ ID 230>. Analysis of this protein sequence reveals the following:

```
Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2451(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 133/222 (59%), Positives = 164/222 (72%)

Query: 2   MNLQENKTAIFDNVSKLALKAGRAHESVHIVAVTKYVNCQTTEALIRTGVNHIGENRVDK 61
           M+L  NK  IF+ +      A R ++SV ++AVTKYV+       LI  G+ HI ENRVDK
Sbjct: 1   MDLLTNKKKIFETIRLSTEAANRTNDSVSVIAVTKYVDSTIAGQLIEAGIEHIAENRVDK 60

Query: 62  FLEKYQALKDEKLTWHLIGSLQRRKVKDVINYVDYFHALDSVKLAAEIQKHAQKLIKCFL 121
           FLEKY ALK   + WHLIG+LQRRKVK+VINYVDYFHALDSV+LA EI K A   +KCFL
Sbjct: 61  FLEKYDALKYMPVKWHLIGTLQRRKVKEVINYVDYFHALDSVRLALEINKRADHPVKCFL 120

Query: 122 QVNISREDSKHGFTIEQIDDALNLISRYDKIELIGIMTMAPLKATKEEISSIFEETESLR 181
           QVNIS+E+SKHGF I +ID+A+  I + +KI+L+G+MTMAP  A+KE I +IF +   LR
Sbjct: 121 QVNISKEESKHGFNISEIDEAIGEIGKMEKIQLVGLMTMAPANASKESIITIFRQANQLR 180

Query: 182 KRLQARNIERMPFTELSMGMSRDYDIAIQNGSTFVRIGTSFF 223
           K LQ +  + MPFTELSMGMS DY IAIQ GSTF+RIG +FF
Sbjct: 181 KNLQLKKRKNMPFTELSMGMSNDYPIAIQEGSTFIRIGRAFF 222
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 71

A DNA sequence (GBSx0071) was identified in *S.agalactiae* <SEQ ID 231> which encodes the amino acid sequence <SEQ ID 232>. This protein is predicted to be YlmF. Analysis of this protein sequence reveals the following:

```
Possible site: 58

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2194(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9617> which encodes amino acid sequence <SEQ ID 9618> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC95442 GB:AF068901 YlmF [Streptococcus pneumoniae]
 Identities = 86/200 (43%), Positives = 120/200 (60%), Gaps = 25/200 (12%)

Query: 5   MALKDRFDKIISYFDTDDVSENEVHEVQERTSVQRDSRAATAQEASQRSHMTNSAEEEMI 64
           M+LKDRFD+ I YF T+D   +  +E       +RD    T+  +SQ   +  +   +
Sbjct: 1   MSLKDRFDRFIDYF-TEDEDSSLPYE-------KRDEPVFTSVNSSQEPALPMNQPSQSA 52

Query: 65  GSRPRTYTYDPNRQERQRVQRDNAYQQATPRVQNKDSVRQQREQVTIALKYPRKYEDAQE 124
           G++    T    RQ+    +  N  Q+AT             ++V I ++YPRKYEDA E
Sbjct: 53  GTKENNITRLHARQQ----ELANQSQRAT-------------DKVIIDVRYPRKYEDATE 95

Query: 125 IVDLLIVNECVLIDFQYMLDAQARRCLDYIDGASRVLYGSLQKVGSSMFLLTPANVMVDI 184
           IVDLL  NE +LIDFQYM + QARRCLDY+DGA  VL G+L+KV S+M+LLTP NV+V++
Sbjct: 96  IVDLLAGNESILIDFQYMTEVQARRCLDYLDGACHVLAGNLKKVASTMYLLTPVNVIVNV 155

Query: 185 EEMNIPKTGQETSFDFDMKR 204
           E++ +P   Q+  F FDMKR
Sbjct: 156 EDIRLPDEDQQGEFGFDMKR 175
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 233> which encodes the amino acid sequence <SEQ ID 234>. Analysis of this protein sequence reveals the following:

```
     Possible site: 49
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.64   Transmembrane  142 - 158 ( 142 - 158)

----- Final Results -----
               bacterial membrane --- Certainty=0.1256(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC95442 GB:AF068901 YlmF [Streptococcus pneumoniae]
 Identities = 82/219 (37%), Positives = 113/219 (51%), Gaps = 46/219 (21%)

Query: 5   MAFKDTFNKMISYFDTDEVNEVEEDVAASTDNVIP--RSQQSVRASSHPKQEPRNNHVQQ 62
           M+ KD F++ I YF  DE            D+ +P  +  + V  S +  QEP     Q
Sbjct: 1   MSLKDRFDRFIDYFTEDE------------DSSLPYEKRDEPVFTSVNSSQEPALPMNQP 48

Query: 63  DHQARSQEQTRSQMHPKHGTSERYYQQSQPKEGHEMVDRRKRMSTSSIANRREQYQQSTC 122
              A ++E   +++H +                                +AN     Q
Sbjct: 49  SQSAGTKENNITRLHARQ----------------------------QELAN-----QSQRA 76

Query: 123 SDQTTIALKYPRKYEDAQEIVDLLIVNECVLIDFQFMLDAQARRCLDFIDGASKVLYGSL 182
           +D+  I ++YPRKYEDA EIVDLL  NE +LIDFQ+M + QARRCLD++DGA  VL G+L
Sbjct: 77  TDKVIIDVRYPRKYEDATEIVDLLAGNESILIDFQYMTEVQARRCLDYLDGACHVLAGNL 136

Query: 183 QKVGSSMYLLAPSNVSVNIEEMTIPHTTQDIGFDFDMKR 221
           +KV S+MYLL P NV VN+E++ +P   Q   F FDMKR
Sbjct: 137 KKVASTMYLLTPVNVIVNVEDIRLPDEDQQGEFGFDMKR 175
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 118/222 (53%), Positives = 145/222 (65%), Gaps = 17/222 (7%)

Query: 1   MEGNMALKDRFDKIISYFDTDDVSENEVHEVQERTSV----QRDSRAATAQEAS------ 50
           ME  MA KD F+K+ISYFDTD+V+E E         +V    Q+  RA++  +
Sbjct: 1   MENKMAFKDTFNKMISYFDTDEVNEVEEDVAASTDNVIPRSQQSVRASSHPKQEPRNNHV 60

Query: 51  QRSHMTNSAEEEMIGSRPRTYTYDPNRQERQRVQR----DNAYQQATPRVQNKDSVRQQR 106
```

```
           Q+ H   S E+      P+  T +   Q+ Q  +     D   + +T  + N+    QQ
Sbjct: 61  QQDHQARSQEQTRSQMHPKHGTSERYYQQSQPKEGHEMVDRRKRMSTSSIANRREQYQQS 120

Query: 107 ---EQVTIALKYPRKYEDAQEIVDLLIVNECVLIDFQYMLDAQARRCLDYIDGASRVLYG 163
              +Q TIALKYPRKYEDAQEIVDLLIVNECVLIDFQ+MLDAQARRCLD+IDGAS+VLYG
Sbjct: 121 TCSDQTTIALKYPRKYEDAQEIVDLLIVNECVLIDFQFMLDAQARRCLDFIDGASKVLYG 180

Query: 164 SLQKVGSSMFLLTPANVMVDIEEMNIPKTGQETSFDFDMKRR 205
           SLQKVGSSM+LL P+NV V+IEEM IP T Q+  FDFDMKRR
Sbjct: 181 SLQKVGSSMYLLAPSNVSVNIEEMTIPHTTQDIGFDFDMKRR 222
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 72

A DNA sequence (GBSx0072) was identified in *S.agalactiae* <SEQ ID 235> which encodes the amino acid sequence <SEQ ID 236>. This protein is predicted to be YlmH. Analysis of this protein sequence reveals the following:

```
Possible site: 35

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3956(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC95444 GB:AF068901 YlmH [Streptococcus pneumoniae]
 Identities = 101/255 (39%), Positives = 161/255 (62%)

Query: 6   IYQHFRPEEYAFIHKIDHLAQYVENTYSFITTEFLNPREFKILESVLERRGSHYYTSGQY 65
           IYQHF  E+  F+ K     + VE++Y+   T F+NP + K+L+ + +  G    +SG++
Sbjct: 5   IYQHFSIEDRPFLDKGMEWIKKVEDSYAPFLTPFINPHQEKLLKILAKTYGLACSSSGEF 64

Query: 66  FQTEYVKVIIAPEYYQLDMADFNLSLIEIKYNAKFNHLTHAKIMGTLLNYLGVKRSILGD 125
             +EYV+V++ P+Y+Q + +DF +SL EI Y+ KF HLTHAKI+GT++N LG++R + GD
Sbjct: 65  VSSEYVRVLLYPDYFQPEFSDFEISLQEIVYSNKFEHLTHAKILGTVINQLGIERKLFGD 124

Query: 126 ILVEEGCAQVLVDSQMTNHLVHSVTKIGTASVQLAEVPLSKLLTPKQDIQKLTVIASSLR 185
           ILV+E  AQ++++ Q        + KIG   V L E P ++ +   +  ++L +  SS R
Sbjct: 125 ILVDEERAQIMINQQFLLLFQDGLKKIGRIPVSLEERPFTEKIDKLEQYRELDLSVSSFR 184

Query: 186 LDKILATILKISRTQSTKLIEADKVKVNYATVNRVSEQLVEGDLISVRGYGRFTLNHNLG 245
           LD +L+ +LK+SR Q+ +LIE   V+VNY  V++    +  GDLISVR +GR  L  + G
Sbjct: 185 LDVLLSNVLKLSRNQANQLIEKKLVQVNYHVVDKSDYTVQVGDLISVRKFGRLRLLQDKG 244

Query: 246 LTKNQKYKLEVDKMI 260
            TK +K K+ V  ++
Sbjct: 245 QTKKEKKKITVQLLL 259
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 237> which encodes the amino acid sequence <SEQ ID 238>. Analysis of this protein sequence reveals the following:

```
     Possible site: 56
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.69   Transmembrane   46 -  62 (  46 -  62)

----- Final Results -----
              bacterial membrane --- Certainty=0.1277(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

```
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC95444 GB:AF068901 YlmH [Streptococcus pneumoniae]
 Identities = 110/257 (42%), Positives = 161/257 (61%)

Query: 7   IYQHFHQEEYPFIDRMSDMINRVEDYYLLEVTEFLNPREVMILKSLIALTDLKMFVSTDY 66
           IYQHF  E+ PF+D+  + I +VED Y   +T F+NP +  +LK L     L    S ++
Sbjct: 5   IYQHFSIEDRPFLDKGMEWIKKVEDSYAPFLTPFINPHQEKLLKILAKTYGLACSSSGEF 64

Query: 67  YPSEYGRVIIAPGYYDLEQSDFQIALVEISYQAKFNQLTHSQILGTLINELGVKRNLFGD 126
             SEY RV++ P Y+  E SDF+I+L EI Y  KF  LTH++ILGT+IN+LG++R LFGD
Sbjct: 65  VSSEYVRVLLYPDYFQPEFSDFEISLQEIVYSNKFEHLTHAKILGTVINQLGIERKLFGD 124

Query: 127 VFVEMGYAQLMIKRELLDYFLGTITKIAKTSVKLREVNFDQLIRSIDNSQTLDILVSSFR 186
           + V+   AQ+MI ++ L  F   + KI +  V L E  F + I  ++  + LD+ VSSFR
Sbjct: 125 ILVDEERAQIMINQQFLLLFQDGLKKIGRIPVSLEERPFTEKIDKLEQYRELDLSVSSFR 184

Query: 187 LDGVVATILKKSRTQVIALIEANKIKVNYRVANKASDNLVIGDMVSIRGHGRFTLLADNG 246
           LD +++ +LK SR Q   LIE   ++VNY V +K+   + +GD++S+R  GR  LL D G
Sbjct: 185 LDVLLSNVLKLSRNQANQLIEKKLVQVNYHVVDKSDYTVQVGDLISVRKFGRLRLLQDKG 244

Query: 247 VTKHGKQKITLSKMIHK 263
            TK  K+KIT+  ++ K
Sbjct: 245 QTKKEKKKITVQLLLSK 261
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 123/256 (48%), Positives = 177/256 (69%)

Query: 6   IYQHFRPEEYAFIHKIDHLAQYVENTYSFITTEFLNPREFKILESVLERRGSHYYTSGQY 65
           IYQHF  EEY FI ++  +   VE+ Y    TEFLNPRE  IL+S++       + S  Y
Sbjct: 7   IYQHFHQEEYPFIDRMSDMINRVEDYYLLEVTEFLNPREVMILKSLIALTDLKMFVSTDY 66

Query: 66  FQTEYVKVIIAPEYYQLDMADFNLSLIEIKYNAKFNHLTHAKIMGTLLNYLGVKRSILGD 125
           + +EY +VIIAP YY L+ +DF ++L+EI Y AKFN LTH++I+GTL+N LGVKR++ GD
Sbjct: 67  YPSEYGRVIIAPGYYDLEQSDFQIALVEISYQAKFNQLTHSQILGTLINELGVKRNLFGD 126

Query: 126 ILVEEGCAQVLVDSQMTNHLVHSVTKIGTASVQLAEVPLSKLLTPKQDIQKLTVIASSLR 185
           + VE G AQ+++  ++ ++ + ++TKI   SV+L EV   +L+    + Q L ++ SS R
Sbjct: 127 VFVEMGYAQLMIKRELLDYFLGTITKIAKTSVKLREVNFDQLIRSIDNSQTLDILVSSFR 186

Query: 186 LDKILATILKISRTQSTKLIEADKVKVNYATVNRVSEQLVEGDLISVRGYGRFTLNHNLG 245
           LD ++ATILK SRTQ   LIEA+K+KVNY   N+ S+ LV GD++S+RG+GRFTL  + G
Sbjct: 187 LDGVVATILKKSRTQVIALIEANKIKVNYRVANKASDNLVIGDMVSIRGHGRFTLLADNG 246

Query: 246 LTKNQKYKLEVDKMIH 261
           +TK+ K K+ + KMIH
Sbjct: 247 VTKHGKQKITLSKMIH 262
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 73

A DNA sequence (GBSx0073) was identified in *S.agalactiae* <SEQ ID 239> which encodes the amino acid sequence <SEQ ID 240>. This protein is predicted to be cell division protein DivIVA (septumplacement). Analysis of this protein sequence reveals the following:

```
Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
bacterial cytoplasm --- Certainty=0.5418(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC95445 GB:AF068901 cell division protein DivIVA [Streptococcus pneumoniae]
 Identities = 132/227 (58%), Positives = 179/227 (78%), Gaps = 2/227 (0%)

Query: 1   MPLTALEIKDKTFSSKFRGYSEEEVNEFLEIVVDDYEDLIRRNREQEQYIKDLEEKIAYF 60
            MP+T+LEIKDKTF ++FRG+  EEV+EFL+IVV DYEDL+R N ++    IK LEE+++YF
Sbjct: 1   MPITSLEIKDKTFGTRFRGFDPEEVDEFLDIVVRDYEDLVRANHDKNLRIKSLEERLSYF 60

Query: 61  NEMKESLSQSVILAQETAERVKISAQDEASNLMGKATFDAQHLIDEAKLKANQILRDATD 120
            +E+K+SLSQSV++AQ+TAERVK +A + ++N++ +A  DAQ L++EAK KAN+ILR ATD
Sbjct: 61  DEIKDSLSQSVLIAQDTAERVKQAAHERSNNIIHQAEQDAQRLLEEAKYKANEILRQATD 120

Query: 121 DAKRVAIETEDLKRQSRVFHQRLLSELEGQLKLANSSAWEELLKPTAIYLQNSDASFKEV 180
            +AK+VA+ETE+LK +SRVFHQRL S +E QL +  SS WE++L+PTA YLQ SD +FKEV
Sbjct: 121 NAKKVAVETEELKNKSRVFHQRLKSTIESQLAIVESSDWEDILRPTATYLQTSDEAFKEV 180

Query: 181 VEKVLDEDDALPVVDDTESFDATRQFSPDEMEELQRRVEESNKQLEE 227
            V +VL E     P+  + E  D TRQFS  EM ELQ R+E ++K+L E
Sbjct: 181 VSEVLGEPIPAPI--EEEPIDMTRQFSQAEMAELQARIEVADKELSE 225
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 241> which encodes the amino acid sequence <SEQ ID 242>. Analysis of this protein sequence reveals the following:

```
Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.6272(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 180/254 (70%), Positives = 217/254 (84%), Gaps = 2/254 (0%)

Query: 1   MPLTALEIKDKTFSSKFRGYSEEEVNEFLEIVVDDYEDLIRRNREQEQYIKDLEEKIAYF 60
            M LT LEIKDKTF +KFRGY EEEVNEFL+IVVDDYE L+R+NR+ E  IKDLEEK++YF
Sbjct: 1   MALTTLEIKDKTFKTKFRGYCEEEVNEFLDIVVDDYEALVRKNRDNEARIKDLEEKLSYF 60

Query: 61  NEMKESLSQSVILAQETAERVKISAQDEASNLMGKATFDAQHLIDEAKLKANQILRDATD 120
            +EMKESLSQSVILAQETAE+VK +A  EA+NL+ KAT+DAQHL+DE+K KANQ+LRDATD
Sbjct: 61  DEMKESLSQSVILAQETAEKVKATANAEATNLVSKATYDAQHLLDESKAKANQMLRDATD 120

Query: 121 DAKRVAIETEDLKRQSRVFHQRLLSELEGQLKLANSSAWEELLKPTAIYLQNSDASFKEV 180
            +AKRVAIETE+LKRQ+RVFHQRL+S +E QL L+NS  W+ELL+PTAIYLQNSD +FKEV
Sbjct: 121 EAKRVAIETEELKRQTRVFHQRLISSIESQLSLSNSPEWDELLQPTAIYLQNSDDAFKEV 180

Query: 181 VEKVLDEDDALPVVDDTESFDATRQFSPDEMEELQRRVEESNKQLEESGLLDTNNFQMEE 240
            V+ VL+ED  +P  DD+ SFDATRQF+P+E+EELQRRV+ESNK+LE   L   ++   E
Sbjct: 181 VKTVLNED--IPESDDSASFDATRQFTPEELEELQRRVDESNKELEAYQLDSQSDSTTEP 238

Query: 241 PINLGETQTFKLNI 254
             +NL ETQTFKLNI
Sbjct: 239 EVNLSETQTFKLNI 252
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 74

A DNA sequence (GBSx0074) was identified in *S.agalactiae* <SEQ ID 243> which encodes the amino acid sequence <SEQ ID 244>. Analysis of this protein sequence reveals the following:

```
Possible site: 61

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.43   Transmembrane  841 - 857 ( 841 - 857)

----- Final Results -----
               bacterial membrane --- Certainty=0.1171(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC95446 GB:AF068901 isoleucine-tRNA synthetase [Streptococcus pneumoniae]
 Identities = 730/929 (78%), Positives = 822/929 (87%), Gaps = 1/929 (0%)

Query: 1   MKLKETLNLGQTAFPMRAGLPNKEPQWQEAWDQADIYKKRQALNEGKPAFHLHDGPPYAN 60
            MKLK+TLNLG+T FPMRAGLP KEP WQ+ W+ A +Y++RQ LN+GKP F LHDGPPYAN
Sbjct: 1   MKLKDTLNLGKTEFPMRAGLPTKEPVWQKEWEDAKLYQRRQELNQGKPHFTLHDGPPYAN 60

Query: 61  GNIHVGHALNKISKDIIVRSKSMSGFRAPYVPGWDTHGLPIEQVLAKKGVKRKEMDLAEY 120
            GNIHVGHA+NKISKDIIVRSKSMSGF AP++PGWDTHGLPIEQVL+K+GVKRKEMDL EY
Sbjct: 61  GNIHVGHAMNKISKDIIVRSKSMSGFYAPFIPGWDTHGLPIEQVLSKQGVKRKEMDLVEY 120

Query: 121 LEMCRDYALSQVDKQRDDFKRLGVSADWENPYITLTPDYEADQVRVFGAMADKGYIYRGA 180
            L++CR+YALSQVDKQR+DFKRLGVS DWENPY+TLTPDYEA Q+RVFG MA+KGYIYRGA
Sbjct: 121 LKLCREYALSQVDKQREDFKRLGVSGDWENPYVTLTPDYEAAQIRVFGEMANKGYIYRGA 180

Query: 181 KPVYWSWSSESALAEAEIEYHDIDSTSLYYANKVKDGKGILDTDTYIVVWTTTPFTVTAS 240
            KPVYWSWSSESALAEAEIEYHD+ STSLYYANKVKDGKG+LDTDTYIVVWTTTPFT+TAS
Sbjct: 181 KPVYWSWSSESALAEAEIEYHDLVSTSLYYANKVKDGKGVLDTDTYIVVWTTTPFTITAS 240

Query: 241 RGLTVGPDMEYVVVVPVGSERKYLLAEVLVDSLAAKFGWENFEIVTHHTGKELNHIVTEH 300
            RGLTVG D++YV+V PVG  RK+++A  L+ SL+ KFGW + +++   + G+ELNHIVTEH
Sbjct: 241 RGLTVGADIDYVLVQPVGEARKFVVAAELLTSLSEKFGWADVQVLETYRGQELNHIVTEH 300

Query: 301 PWDTEVEELVILGDHVTTDSGTGIVHTAPGFGEDDYNVGIANGLDVVVTVDSRGLMMENA 360
            PWDT VEELVILGDHVTTDSGTGIVHTAPGFGEDDYNVGIAN L+V VTVD RG+MM+NA
Sbjct: 301 PWDTAVEELVILGDHVTTDSGTGIVHTAPGFGEDDYNVGIANNLEVAVTVDERGIMMKNA 360

Query: 361 GPDFEGQFYDKVTPLVKEKLGDLLLASEVINHSYPFDWRTKKPIIWRAVPQWFASVSKFR 420
            GP+FEGQFY+KV P V EKLG+LLLA E I+HSYPFDWRTKKPIIWRAVPQWFASVSKFR
Sbjct: 361 GPEFEGQFYEKVVPTVIEKLGNLLLAQEEISHSYPFDWRTKKPIIWRAVPQWFASVSKFR 420

Query: 421 QEILDEIEKTNFQPEWGKKRLYNMIRDRGDWVISRQRAWGVPLPIFYAEDGTAIMTKEVT 480
            QEILDEIEK  F  EWGK RLYNMIRDRGDWVISRQR WGVPLPIFYAEDGTAIM  E
Sbjct: 421 QEILDEIEKVKFHSEWGKVRLYNMIRDRGDWVISRQRTWGVPLPIFYAEDGTAIMVAETI 480

Query: 481 DHVADLFAEYGSIVWWQRDAKDLLPAGYTHPGSPNGLFEKETDIMDVWFDSGSSWNGVMN 540
            +HVA LF ++GS +WW+RDAKDLLP G+THPGSPNG F+KETDIMDVWFDSGSSWNGV+
Sbjct: 481 EHVAQLFEKHGSSIWWERDAKDLLPEGFTHPGSPNGEFKKETDIMDVWFDSGSSWNGVVV 540

Query: 541 ARENLSYPADLYLEGSDQYRGWFNSSLITSVAVNGHAPYKAVLSQGFVLDGKGEKMSKSL 600
             R  L+YPADLYLEGSDQYRGWFNSSLITSVA +G APYK +LSQGF LDGKGEKMSKSL
Sbjct: 541 NRPELTYPADLYLEGSDQYRGWFNSSLITSVANHGVAPYKQILSQGFALDGKGEKMSKSL 600

Query: 601 GNTILPSDVEKQFGAEILRLWVTSVDSSNDVRISMDILKQTSETYRKIRNTLRFLIANTS 660
            GNTI PSDVEKQFGAEILRLWVTSVDSSNDVRISMDIL Q SETYRKIRNTLRFLIANTS
Sbjct: 601 GNTIAPSDVEKQFGAEILRLWVTSVDSSNDVRISMDILSQVSETYRKIRNTLRFLIANTS 660

Query: 661 DFNPKQDAVAYENLGAVDRYMTIKFNQVVDTINKAYAAYDFMAIYKAVVNFVTVDLSAFY 720
            DFNP QD VAY+ L +VD+YMTI+FNQ+V TI  AYA ++F+ IYKA+VNF+ VDLSAFY
Sbjct: 661 DFNPAQDTVAYDELRSVDKYMTIRFNQLVKTIRDAYADFEFLTIYKALVNFINVDLSAFY 720
```

```
Query: 721 LDFAKDVVYIEAANSPERRRMQTVFYDILVKLTKLLTPILPHTAEEIWSYLEHEEEEFVQ 780
           LDFAKDVVYIE A S ERR+MQTVFYDILVK+TKLLTPILPHTAEEIWSYLE E E+FVQ
Sbjct: 721 LDFAKDVVYIEGAKSLERRQMQTVFYDILVKITKLLTPILPHTAEEIWSYLEFETEDFVQ 780

Query: 781 LAEMPVAQTFSGQEEILEEWSAFMTLRTQAQKALEEARNAKVIGKSLEAHLTIYASQEVK 840
           L+E+P  QTF+ QEEIL+ W+AFM  R QAQKALEEARNAKVIGKSLEAHLT+Y ++ VK
Sbjct: 781 LSELPEVQTFANQEEILDTWAAFMDFRGQAQKALEEARNAKVIGKSLEAHLTVYPNEVVK 840

Query: 841 TLLTALNSDIALLMIVSQLTIADEADKPADSVSFEGVAFTVEHAEGEVCERSRRIDPTTK 900
           TLL A+NS++A L+IVS+LTIA+E   P  ++SFE VAFTVE A GEVC+R RRIDPTT
Sbjct: 841 TLLEAVNSNVAQLLIVSELTIAEE-PAPEAALSFEDVAFTVERAAGEVCDRCRRIDPTTA 899

Query: 901 MRSYGVAVCDASAAIIEQYYPEAVAQGFE 929
            RSY   +CD  A+I+E+ + +AVA+GFE
Sbjct: 900 ERSYQAVICDHCASIVEENFADAVAEGFE 928
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 245> which encodes the amino acid sequence <SEQ ID 246>. Analysis of this protein sequence reveals the following:

```
Possible site: 61

>>> Seems to have no N-terminal signal sequence
   INTEGRAL     Likelihood = -1.70   Transmembrane  849 - 865 ( 848 - 867)

----- Final Results -----
               bacterial membrane --- Certainty=0.1680(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 798/929 (85%), Positives = 857/929 (91%)

Query: 1   MKLKETLNLGQTAFPMRAGLPNKEPQWQEAWDQADIYKKRQALNEGKPAFHLHDGPPYAN 60
           MKLKETLNLG+TAFPMRAGLPNKEPQWQ AW+QA++YKKRQ LN GKPAFHLHDGPPYAN
Sbjct: 1   MKLKETLNLGKTAFPMRAGLPNKEPQWQAAWEQAELYKKRQELNAGKPAFHLHDGPPYAN 60

Query: 61  GNIHVGHALNKISKDIIVRSKSMSGFRAPYVPGWDTHGLPIEQVLAKKGVKRKEMDLAEY 120
           GNIHVGHALNKISKDIIVRSKSMSGF+APYVPGWDTHGLPIEQVLAK+G+KRKEMDLAEY
Sbjct: 61  GNIHVGHALNKISKDIIVRSKSMSGFQAPYVPGWDTHGLPIEQVLAKQGIKRKEMDLAEY 120

Query: 121 LEMCRDYALSQVDKQRDDFKRLGVSADWENPYITLTPDYEADQVRVFGAMADKGYIYRGA 180
           LEMCR YALSQVDKQRDDFKRLGVSADWENPY+TL P +EADQ+RVFGAMA+KGYIYRGA
Sbjct: 121 LEMCRQYALSQVDKQRDDFKRLGVSADWENPYVTLDPQFEADQIRVFGAMAEKGYIYRGA 180

Query: 181 KPVYWSWSSESALAEAEIEYHDIDSTSLYYANKVKDGKGILDTDTYIVVWTTTPFTVTAS 240
           KPVYWSWSSESALAEAEIEYHDIDSTSLYYANKVKDGKGILDT+TYIVVWTTTPFTVTAS
Sbjct: 181 KPVYWSWSSESALAEAEIEYHDIDSTSLYYANKVKDGKGILDTNTYIVVWTTTPFTVTAS 240

Query: 241 RGLTVGPDMEYVVVVPVGSERKYLLAEVLVDSLAAKFGWENFEIVTHHTGKELNHIVTEH 300
           RGLTVGPDM+Y+VV P GS+R+Y++AE L+DSLA KFGWE+FE +  H G +L +IVTEH
Sbjct: 241 RGLTVGPDMDYLVVKPAGSDRQYVVAEGLLDSLAGKFGWESFETLASHKGADLEYIVTEH 300

Query: 301 PWDTEVEELVILGDHVTTDSGTGIVHTAPGFGEDDYNVGIANGLDVVVTVDSRGLMMENA 360
           PWDT+VEELVILGDHVT +SGTGIVHTAPGFGEDDYNVG    L+V VTVD RGLMMENA
Sbjct: 301 PWDTDVEELVILGDHVTLESGTGIVHTAPGFGEDDYNVGTKYKLEVAVTVDERGLMMENA 360

Query: 361 GPDFEGQFYDKVTPLVKEKLGDLLLASEVINHSYPFDWRTKKPIIWRAVPQWFASVSKFR 420
           GPDF GQFY+KVTP+V +KLGDLLLA EVINHSYPFDWRTKKPIIWRAVPQWFASVS FR
Sbjct: 361 GPDFHGQFYNKVTPIVIDKLGDLLLAQEVINHSYPFDWRTKKPIIWRAVPQWFASVSDFR 420

Query: 421 QEILDEIEKTNFQPEWGKKRLYNMIRDRGDWVISRQRAWGVPLPIFYAEDGTAIMTKEVT 480
           Q+ILDEIEKT F P WG+ RLYNMIRDRGDWVISRQRAWGVPLPIFYAEDGTAIMTKEVT
Sbjct: 421 QDILDEIEKTTFHPSWGETRLYNMIRDRGDWVISRQRAWGVPLPIFYAEDGTAIMTKEVT 480

Query: 481 DHVADLFAEYGSIVWWQRDAKDLLPAGYTHPGSPNGLFEKETDIMDVWFDSGSSWNGVMN 540
```

```
            DHVADLF E GSI+WWQ++AKDLLP G+THPGSPNG F KETDIMDVWFDSGSSWNGVMN
Sbjct: 481 DHVADLFQENGSIIWWQKEAKDLLPEGFTHPGSPNGEFTKETDIMDVWFDSGSSWNGVMN 540

Query: 541 ARENLSYPADLYLEGSDQYRGWFNSSLITSVAVNGHAPYKAVLSQGFVLDGKGEKMSKSL 600
            +ENLSYPADLYLEGSDQYRGWFNSSLITSVAVNGHAPYKA+LSQGFVLDGKGEKMSKS
Sbjct: 541 TKENLSYPADLYLEGSDQYRGWFNSSLITSVAVNGHAPYKAILSQGFVLDGKGEKMSKSK 600

Query: 601 GNTILPSDVEKQFGAEILRLWVTSVDSSNDVRISMDILKQTSETYRKIRNTLRFLIANTS 660
            GN I P+DV KQ+GA+ILRLWV SVD+ NDVR+SM+IL Q SETYRKIRNTLRFLIANTS
Sbjct: 601 GNIISPNDVAKQYGADILRLWVASVDTDNDVRVSMEILGQVSETYRKIRNTLRFLIANTS 660

Query: 661 DFNPKQDAVAYENLGAVDRYMTIKFNQVVDTINKAYAAYDFMAIYKAVVNFVTVDLSAFY 720
            DFNP  D VAY +LG VD+YMTI FNQ+V TI  AY  YDFMAIYKAVVNFVTVDLSAFY
Sbjct: 661 DFNPATDTVAYADLGTVDKYMTIVFNQLVATITDAYERYDFMAIYKAVVNFVTVDLSAFY 720

Query: 721 LDFAKDVVYIEAANSPERRRMQTVFYDILVKLTKLLTPILPHTAEEIWSYLEHEEEEFVQ 780
            LDFAKDVVYIEAANS ERRRMQTVFYDILVK+TKLLTPILPHT EEIWSYLEHE E FVQ
Sbjct: 721 LDFAKDVVYIEAANSLERRRMQTVFYDILVKITKLLTPILPHTTEEIWSYLEHESEAFVQ 780

Query: 781 LAEMPVAQTFSGQEEILEEWSAFMTLRTQAQKALEEARNAKVIGKSLEAHLTIYASQEVK 840
            LAEMPVA+TFS QE+ILE WSAFMTLRTQAQKALEEARNAK+IGKSLEAHLTIYAS+EVK
Sbjct: 781 LAEMPVAETFSAQEDILEAWSAFMTLRTQAQKALEEARNAKIIGKSLEAHLTIYASEEVK 840

Query: 841 TLLTALNSDIALLMIVSQLTIADEADKPADSVSFEGVAFTVEHAEGEVCERSRRIDPTTK 900
            TLLTAL+SDIALL+IVSQLTIAD AD PAD+V+FEGVAF VEHA GEVCERSRRIDPTT+
Sbjct: 841 TLLTALDSDIALLLIVSQLTIADLADAPADAVAFEGVAFIVEHAIGEVCERSRRIDPTTR 900

Query: 901 MRSYGVAVCDASAAIIEQYYPEAVAQGFE 929
            MRSY   VCD SA IIE+ +PEAVA+GFE
Sbjct: 901 MRSYNAFVCDHSAKIIEENFPEAVAEGFE 929
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 75

A DNA sequence (GBSx0075) was identified in *S.agalactiae* <SEQ ID 247> which encodes the amino acid sequence <SEQ ID 248>. Analysis of this protein sequence reveals the following:

```
Possible site: 39

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3425(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 249> which encodes the amino acid sequence <SEQ ID 250>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3467(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 77/99 (77%), Positives = 89/99 (89%)
```

```
Query: 1  MRLINTTSSHPELVRNQLQNTDAKLVEVYSAGNTDVVFTKAPKHYELLISNKYRAIKDEE 60
          MRLINTTSSHPEL++NQL+NTDA LVEVYSAGNTDV+FT+APKHYELLISNKYRAIK++E
Sbjct: 1  MRLINTTSSHPELIKNQLKNTDAYLVEVYSAGNTDVIFTQAPKHYELLISNKYRAIKEDE 60

Query: 61 LEAIREFFLKRKIDQSIIIQEQMKSLHTAKLIEISYPTT 99
          L+ IREFFLKRKID  I+I  Q K+LHT  LIEIS+ T+
Sbjct: 61 LDIIREFFLKRKIDPKIVIPGQSKTLHTNNLIEISFQTS 99
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 76

A DNA sequence (GBSx0076) was identified in *S.agalactiae* <SEQ ID 251> which encodes the amino acid sequence <SEQ ID 252>. This protein is predicted to be AP4A hydrolase. Analysis of this protein sequence reveals the following:

```
Possible site: 42

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1714(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC06510 GB:AE000676 AP4A hydrolase [Aquifex aeolicus]
 Identities = 30/101 (29%), Positives = 48/101 (46%), Gaps = 2/101 (1%)

Query: 32  KIILVQAPNGAWFLPGGEIEENENHLEALTRELIEELGYSATIGHYYGQADEYFYSRHRD 91
           +++L++ P+  W  P G IE  E   E   RE+ EE G    I  Y G+   Y+Y+   +
Sbjct: 16  EVLLIKTPSNVWSFPKGNIEPGEKPEETAVREVWEETGVKGEILDYIGEI-HYWYTLKGE 74

Query: 92  TYYYNPAYIYEVTAYHKDQAPLEDFNHLAWFPIQEAKEKLK 132
             +    Y Y +     +  P  +     +FPI+EAK+ LK
Sbjct: 75  RIFKTVKY-YLMKYKEGEPRPSWEVKDAKFFPIKEAKKLLK 114
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 253> which encodes the amino acid sequence <SEQ ID 254>. Analysis of this protein sequence reveals the following:

```
Possible site: 47

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1954(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 102/149 (68%), Positives = 118/149 (78%)

Query: 1   MTNPTFGEKIDNVNYRSRFGVYAIIPNPTHDKIILVQAPNGAWFLPGGEIEENENHLEAL 60
           M  PTFG K  + +Y +R+GVYAIIPN    KIILVQAPNG+WFLPGGEIE  E  L+AL
Sbjct: 1   MMIPTFGHKNAHKDYVTRYGVYAIIPNHEQTKIILVQAPNGSWFLPGGEIEAGEGQLQAL 60

Query: 61  TRELIEELGYSATIGHYYGQADEYFYSRHRDTYYYNPAYIYEVTAYHKDQAPLEDFNHLA 120
            RELIEELG+SATIG YYGQADEYFYSRHRDT++Y+PAY+YEVTA+     PLEDFN+L
Sbjct: 61  ERELIEELGFSATIGSYYGQADEYFYSRHRDTHFYHPAYLYEVTAFQAVSKPLEDFNNLG 120
```

```
Query: 121 WFPIQEAKEKLKRGSHRWGVQAWEKNHHS 149
           WF   EA  KLKR SH+WGV+ W+K HHS
Sbjct: 121 WFSPIEAIAKLKRESHQWGVKEWQKKHHS 149
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 77

A DNA sequence (GBSx0077) was identified in *S.agalactiae* <SEQ ID 255> which encodes the amino acid sequence <SEQ ID 256>. This protein is predicted to be ClpE (clpB-1). Analysis of this protein sequence reveals the following:

```
Possible site: 54

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2882(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD01782 GB:AF023421 ClpE [Lactococcus lactis]
 Identities = 560/752 (74%), Positives = 647/752 (85%), Gaps = 12/752 (1%)

Query: 1   MLCQNCKLNESTIHLYTNVNGKQKQVDLCQNCYQIIKTDPNNPLFSGLNHVS-HAPGGIN 59
           MLCQNC +NE+TIHLYT+VNG++KQ+DLCQNCYQI+K+      LF   N  + ++    N
Sbjct: 1   MLCQNCNINEATIHLYTSVNGQKKQIDLCQNCYQIMKSGGQEALFGAGNASNGNSDEPFN 60

Query: 60  PFFDDFFGDLNNFRAFNGQDLPNTPPTQSGGNRGGGNGNGRNNNRNQTATPSQAKGILEE 119
           PF +D F  L   + FNG     TPPTQ+GG    G  N R           Q KG+LEE
Sbjct: 61  PF-NDIFSALQG-QDFNGAASNQTPPTQTGGRGPRGPQNPR---------AKQPKGMLEE 109

Query: 120 FGINVTEIARHGDIDPVIGRDSEIIRVIEILNRRTKNNPVLIGEPGVGKTAVVEGLAQKI 179
           FGIN+TE AR G+IDPVIGRD EI RVIEILNRRTKNNPVLIGEPGVGKTAVVEGLAQKI
Sbjct: 110 FGINITESARRGEIDPVIGRDEEIKRVIEILNRRTKNNPVLIGEPGVGKTAVVEGLAQKI 169

Query: 180 VDGNVPHKLQGKQVIRLDVVSLVQGTGIRGQFEERMQKLMEEIRQRQDVILFIDEIHEIV 239
           VDG+VP KLQ K+VIRLDVVSLVQGTGIRGQFEERMQKLM+EIR+R DVI+FIDEIHEIV
Sbjct: 170 VDGDVPQKLQNKEVIRLDVVSLVQGTGIRGQFEERMQKLMDEIRKRNDVIMFIDEIHEIV 229

Query: 240 GAGTAGEGSMDAGNILKPALARGELQLVGATTLNEYRIIEKDAALERRMQPVKVDEPSVE 299
           GAG+AG+G+MDAGNILKPALARGELQLVGATTLNEYRIIEKDAALERRMQPVKVDEPSV+
Sbjct: 230 GAGSAGDGNMDAGNILKPALARGELQLVGATTLNEYRIIEKDAALERRMQPVKVDEPSVD 289

Query: 300 ETITILKGIQKKYEDYHHVKYNNDAIEAAAVLSNRYIQDRFLPDKAIDLLDEAGSKMNLT 359
           ETITIL+GIQ +YEDYHHVKY ++AIEAAA LSNRYIQDRFLPDKAIDLLDE+GSK NLT
Sbjct: 290 ETITILRGIQARYEDYHHVKYTDEAIEAAAHLSNRYIQDRFLPDKAIDLLDESGSKKNLT 349

Query: 360 LNFVDPKEIDQRLIEAENLKAQATREEDYERAAYFRDQIAKYKEMQQQKVDDQDTPIITE 419
           L FVDP++I++R+  +AE+ K +AT+ ED+E+AA+FRDQI+K +E+Q+Q+V D+D P+ITE
Sbjct: 350 LKFVDPEDINRRIADAESKKNEATKAEDFEKAAHFRDQISKLRELQKQEVTDEDMPVITE 409

Query: 420 KTIEHIIEEKTNIPVGDLKEKEQSQLINLADDLKQHVIGQDDAVVKIAKAIRRNRVGLGS 479
           K IE I+E+KT IPVGDLKEKEQ+QLINLADDLK HVIGQD+AV KI+KAIRR+RVGLG
Sbjct: 410 KDIEQIVEQKTQIPVGDLKEKEQTQLINLADDLKAHVIGQDEAVDKISKAIRRSRVGLGK 469

Query: 480 PNRPIGSFLFVGPTGVGKTELSKQLAIELFGSADSMIRFDMSEYMEKHAVAKLVGAPPGY 539
           PNRPIG FLFVGPTGVGKTEL+KQLA ELFGS++SMIRFDMSEYMEKH+VAKL+GAPPGY
Sbjct: 470 PNRPIGFFLFVGPTGVGKTELAKQLAKELFGSSESMIRFDMSEYMEKHSVAKLIGAPPGY 529

Query: 540 VGYEEAGQLTEKVRRNPYSLILLDEIEKAHPDVMHMFLQVLDDGRLTDGQGRTVSFKDTI 599
           VGYEEAGQLTE+VRRNPYSLILLDEIEKAHPDVMHMFLQ+L+DGRLTD QGRTVSFKD++
```

```
Sbjct: 530 VGYEEAGQLTERVRRNPYSLILLDEIEKAHPDVMHMFLQILEDGRLTDAQGRTVSFKDSL 589

Query: 600 IIMTSNAGSGKTEASVGFGASREGRTNSVLGQLGNFFSPEFMNRFDGIIEFKALDKENLL 659
           IIMTSNAG+GK EASVGFGA+REGRT SVLGQLG+FFSPEFMNRFDGIIEF AL KENLL
Sbjct: 590 IIMTSNAGTGKVEASVGFGAAREGRTKSVLGQLGDFFSPEFMNRFDGIIEFSALSKENLL 649

Query: 660 NIVDIMLSDVNARLAINGIHLDVTDKVKEKLVDLGYDPKMGARPLRRTIQEHIEDAITDY 719
            IVD+ML +VN ++  N IHL VT   KEKLVDLGY+P MGARPLRR IQE+IED+I D+
Sbjct: 650 KIVDLMLDEVNEQIGRNDIHLSVTQAAKEKLVDLGYNPAMGARPLRRIIQENIEDSIADF 709

Query: 720 YLENPSEKELRAIMTSNGNIIIKSSKKTEEST 751
           Y+E+P  K+L A +  +  +I   +++T E+T
Sbjct: 710 YIEHPEYKQLVADLIDDKIVISNQTQETAETT 741
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 257> which encodes the amino acid sequence <SEQ ID 258>. Analysis of this protein sequence reveals the following:

```
Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3104(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 640/751 (85%), Positives = 691/751 (91%), Gaps = 7/751 (0%)

Query: 1   MLCQNCKLNESTIHLYTNVNGKQKQVDLCQNCYQIIKTDPNNPLFSGLNHVSHAPG-GIN 59
           MLCQNC LNESTIHLYT+VNGKQ+QVDLCQNCYQI+K+DP N + +GL     A     +
Sbjct: 1   MLCQNCNLNESTIHLYTSVNGKQRQVDLCQNCYQIMKSDPANSILNGLTPGYRAQDRSTS 60

Query: 60  PFFDDFFGDLNNFRAFNGQDLPNTPPTQSGGNRGGGNGNGRNNNRNQTATPS----QAKG 115
           PFFDDFFGDLNNFRAF   +LPNTPPTQ+G N  GG   G N N  + A P     QAKG
Sbjct: 61  PFFDDFFGDLNNFRAFG--NLPNTPPTQAGQNGNGGGRYGGNYNGQRPAQPQTPNQQAKG 118

Query: 116 ILEEFGINVTEIARHGDIDPVIGRDSEIIRVIEILNRRTKNNPVLIGEPGVGKTAVVEGL 175
           +LEEFGINVT+IAR+G+IDPVIGRD EI RVIEILNRRTKNNPVLIGEPGVGKTAVVEGL
Sbjct: 119 LLEEFGINVTDIARNGNIDPVIGRDEEITRVIEILNRRTKNNPVLIGEPGVGKTAVVEGL 178

Query: 176 AQKIVDGNVPHKLQGKQVIRLDVVSLVQGTGIRGQFEERMQKLMEEIRQRQDVILFIDEI 235
           AQKI+DG VP KLQGKQVIRLDVVSLVQGTGIRGQFEERMQKLMEEIR R+DVILFIDEI
Sbjct: 179 AQKIIDGTVPQKLQGKQVIRLDVVSLVQGTGIRGQFEERMQKLMEEIRNRKDVILFIDEI 238

Query: 236 HEIVGAGTAGEGSMDAGNILKPALARGELQLVGATTLNEYRIIEKDAALERRMQPVKVDE 295
           HEIVGAG+AG+G+MDAGNILKPALARGELQLVGATTLNEYRIIEKDAALERRMQPVKVDE
Sbjct: 239 HEIVGAGSAGDGNMDAGNILKPALARGELQLVGATTLNEYRIIEKDAALERRMQPVKVDE 298

Query: 296 PSVEETITILKGIQKKYEDYHHVKYNNDAIEAAAVLSNRYIQDRFLPDKAIDLLDEAGSK 355
           PSVEETITILKGIQ KYEDYHHVKY+  AIEAAA LSNRYIQDRFLPDKAIDLLDEAGSK
Sbjct: 299 PSVEETITILKGIQPKYEDYHHVKYSPAAIEAAAHLSNRYIQDRFLPDKAIDLLDEAGSK 358

Query: 356 MNLTLNFVDPKEIDQRLIEAENLKAQATREEDYERAAYFRDQIAKYKEMQQQKVDDQDTP 415
           MNLTLNFVDPKEID+RLIEAENLKAQATR+EDYERAAYFRDQI KYKEMQ QKVD+QD P
Sbjct: 359 MNLTLNFVDPKEIDKRLIEAENLKAQATRDEDYERAAYFRDQITKYKEMQAQKVDEQDIP 418

Query: 416 IITEKTIEHIIEEKTNIPVGDLKEKEQSQLINLADDLKQHVIGQDDAVVKIAKAIRRNRV 475
           IITEKTIE I+E+KTNIPVGDLKEKEQSQL+NLA+DLK HVIGQDDAV KIAKAIRRNRV
Sbjct: 419 IITEKTIEAIVEQKTNIPVGDLKEKEQSQLVNLANDLKAHVIGQDDAVDKIAKAIRRNRV 478

Query: 476 GLGSPNRPIGSFLFVGPTGVGKTELSKQLAIELFGSADSMIRFDMSEYMEKHAVAKLVGA 535
           GLG+PNRPIGSFLFVGPTGVGKTELSKQLAIELFGS ++MIRFDMSEYMEKHAVAKLVGA
Sbjct: 479 GLGTPNRPIGSFLFVGPTGVGKTELSKQLAIELFGSTNNMIRFDMSEYMEKHAVAKLVGA 538

Query: 536 PPGYVGYEEAGQLTEKVRRNPYSLILLDEIEKAHPDVMHMFLQVLDDGRLTDGQGRTVSF 595
```

```
            PPGY+GYEEAGQLTE+VRRNPYSLILLDE+EKAHPDVMHMFLQVLDDGRLTDGQGRTVSF
Sbjct: 539 PPGYIGYEEAGQLTEQVRRNPYSLILLDEVEKAHPDVMHMFLQVLDDGRLTDGQGRTVSF 598

Query: 596 KDTIIIMTSNAGSGKTEASVGFGASREGRTNSVLGQLGNFFSPEFMNRFDGIIEFKALDK 655
            KDTIIIMTSNAG+GK+EASVGFGA+REGRT+SVLG+L NFFSPEFMNRFDGIIEFKAL K
Sbjct: 599 KDTIIIMTSNAGTGKSEASVGFGAAREGRTSSVLGELSNFFSPEFMNRFDGIIEFKALSK 658

Query: 656 ENLLNIVDIMLSDVNARLAINGIHLDVTDKVKEKLVDLGYDPKMGARPLRRTIQEHIEDA 715
            E+LL+IVD+ML DVN RL  NGIHLDVT KVKEKLVDLGYDPKMGARPLRRTIQ++IEDA
Sbjct: 659 EHLLHIVDLMLEDVNERLGYNGIHLDVTQKVKEKLVDLGYDPKMGARPLRRTIQDYIEDA 718

Query: 716 ITDYYLENPSEKELRAIMTSNGNIIIKSSKK 746
            ITDYYLE+P+EK+LRA+MT++ NI IK+ K+
Sbjct: 719 ITDYYLEHPTEKQLRALMTNSENITIKAVKE 749
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 78

A DNA sequence (GBSx0078) was identified in *S.agalactiae* <SEQ ID 259> which encodes the amino acid sequence <SEQ ID 260>. This protein is predicted to be glutamine ABC transporter, permease protein (glnP). Analysis of this protein sequence reveals the following:

```
Possible site: 61

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.92   Transmembrane   27 -  43 (  15 -  46)
   INTEGRAL    Likelihood = -2.50   Transmembrane  200 - 216 ( 196 - 217)

----- Final Results -----
               bacterial membrane --- Certainty=0.4970(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9619> which encodes amino acid sequence <SEQ ID 9620> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB91000 GB:AE001090 glutamine ABC transporter, permease protein
            (glnP) [Archaeoglobus fulgidus]
 Identities = 92/209 (44%), Positives = 129/209 (61%), Gaps = 10/209 (4%)

Query: 17  YGVMVTIMISTCVVFFGTIIGVLIALVKRTNLHFLTILANFYVWVFRGTPMVVQIMIAFA 76
           +G  VT+ ++   +FFG IIG +  L + +       ++  YV V RGTP++VQI+I +
Sbjct: 21  FGASVTLKLTLISIFFGLIIGTIAGLGRVSKNPLPFAISTAYVEVIRGTPLLVQILIVYF 80

Query: 77  WMHFNNLPTISFGVLDLDFTRLLPGIIIISLNSGAYISEIVRAGIEAVPSGQIEAAYSLG 136
                 LP I   +          GII +S+ SGAYI+EIVRAGIE++P GQ+EAA SLG
Sbjct: 81  -----GLPAIGINLQPEP-----AGIIALSICSGAYIAEIVRAGIESIPIGQMEAARSLG 130

Query: 137 IRPKNTLRYVILPQAFKNILPALGNEFITIIKDSALLQTIGVMELWNGAQSVVTATYSPV 196
           +     +RYVI PQAF+NILPALGNEFI ++KDS+LL  I ++EL    + +V  T++
Sbjct: 131 MTYLQAMRYVIFPQAFRNILPALGNEFIALLKDSSLLSVISIVELTRVGRQIVNTTFNAW 190

Query: 197 APLLFAAFYYLMLTTILSALLKQMEKYLG 225
             P L  A +YLM+T  LS L+   +K LG
Sbjct: 191 TPFLGVALFYLMMTIPLSRLVAYSQKKLG 219
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 261> which encodes the amino acid sequence <SEQ ID 262>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
```

```
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.08   Transmembrane   25 -  41 (  11 -  44)
   INTEGRAL    Likelihood = -1.91   Transmembrane  202 - 218 ( 201 - 218)

----- Final Results -----
              bacterial membrane --- Certainty=0.4630(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAB91000 GB:AE001090 glutamine ABC transporter, permease protein
          (glnP) [Archaeoglobus fulgidus]
 Identities = 91/209 (43%), Positives = 138/209 (65%), Gaps = 12/209 (5%)

Query: 15  YGVLVTIMISVSVVFFGTLIGVLVTLIKRSHVKPLTWVVNL-YVWIFRGTPMVVQIMIAF 73
           +G  VT+ +++   +FFG +IG +  L + S   PL + ++  YV + RGTP++VQI+I +
Sbjct: 21  FGASVTLKLTLISIFFGLIIGTIAGLGRVSK-NPLPFAISTAYVEVIRGTPLLVQILIVY 79

Query: 74  AWMHFNNMPTIGFGVLDLDFSRLLPGIIIISLNSGAYISEIVRAGIEAVPKGQLEAAYSL 133
                  +P IG     ++      GII +S+ SGAYI+EIVRAGIE++P GQ+EAA SL
Sbjct: 80  F-----GLPAIG-----INLQPEPAGIIALSICSGAYIAEIVRAGIESIPIGQMEAARSL 129

Query: 134 GIRPQNAMRYVILPQAFKNILPALGNEFITIIKDSALLQTIGVMELWNGAQSVVTATYSP 193
           G+    AMRYVI PQAF+NILPALGNEFI ++KDS+LL  I ++EL    + +V  T++
Sbjct: 130 GMTYLQAMRYVIFPQAFRNILPALGNEFIALLKDSSLLSVISIVELTRVGRQIVNTTFNA 189

Query: 194 ISPLLVAAFYYLMVTTVMAQLLAVLERHM 222
            +P L  A +YLM+T  +++L+A  ++ +
Sbjct: 190 WTPFLGVALFYLMMTIPLSRLVAYSQKKL 218
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 180/225 (80%), Positives = 208/225 (92%)

Query: 3   MNFSFLPQYWSYFNYGVMVTIMISTCVVFFGTIIGVLIALVKRTNLHFLTILANFYVWVF 62
           M+ SFLP+YW+YFNYGV+VTIMIS  VVFFGT+IGVL+ L+KR+++  LT + N YVW+F
Sbjct: 1   MDLSFLPKYWAYFNYGVLVTIMISVSVVFFGTLIGVLVTLIKRSHVKPLTWVVNLYVWIF 60

Query: 63  RGTPMVVQIMIAFAWMHFNNLPTISFGVLDLDFTRLLPGIIIISLNSGAYISEIVRAGIE 122
           RGTPMVVQIMIAFAWMHFNN+PTI FGVLDLDF+RLLPGIIIISLNSGAYISEIVRAGIE
Sbjct: 61  RGTPMVVQIMIAFAWMHFNNMPTIGFGVLDLDFSRLLPGIIIISLNSGAYISEIVRAGIE 120

Query: 123 AVPSGQIEAAYSLGIRPKNTLRYVILPQAFKNILPALGNEFITIIKDSALLQTIGVMELW 182
           AVP GQ+EAAYSLGIRP+N +RYVILPQAFKNILPALGNEFITIIKDSALLQTIGVMELW
Sbjct: 121 AVPKGQLEAAYSLGIRPQNAMRYVILPQAFKNILPALGNEFITIIKDSALLQTIGVMELW 180

Query: 183 NGAQSVVTATYSPVAPLLFAAFYYLMLTTILSALLKQMEKYLGKG 227
           NGAQSVVTATYSP++PLL AAFYYLM+TT+++ LL  +E+++ +G
Sbjct: 181 NGAQSVVTATYSPISPLLVAAFYYLMVTTVMAQLLAVLERHMAQG 225
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 79

A DNA sequence (GBSx0079) was identified in *S.agalactiae* <SEQ ID 263> which encodes the amino acid sequence <SEQ ID 264>. This protein is predicted to be phosphomannomutase (manB). Analysis of this protein sequence reveals the following:

```
Possible site: 60

>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
             bacterial cytoplasm --- Certainty=0.5400(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9621> which encodes amino acid sequence <SEQ ID 9622> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB04825 GB:AP001510 phosphomannomutase [Bacillus halodurans]
 Identities = 239/548 (43%), Positives = 344/548 (62%), Gaps = 14/548 (2%)

Query: 4   MNYKEIYQEWLENDSLGKDIKSDLEAIKGDESEIQDRFYKTLEFGTAGLRGKLGAGTNRM 63
           M++++ Y++W   + L  ++K  LEAI GDE +++D FYK LEFGT G+RG++G G NRM
Sbjct: 1   MSWRQRYEKWKGFNELELELKQSLEAIGGDEQQLEDCFYKNLEFGTGGMRGEIGPGPNRM 60

Query: 64  NTYMVGKAAQALANTIIDHGPEAIARGIAVSYDVRYQSKEFAELTCSIMAANGIKSYIYK 123
           NTY + KA++  A  +++ G    A+G+ ++YD R++S EFA      +  +GIK+Y+++
Sbjct: 61  NTYTIRKASEGFARYLLEQGEHVKAQGVVIAYDSRHKSPEFAREAALTIGKHGIKAYLFE 120

Query: 124 GIRPTPMCSYAIRALGCVSGVMITASHNPQAYNGYKAYWKEGSQILDDIADQIANHMDAI 183
            +RPTP  S+A+R LG   G++ITASHNP  YNG+K Y  +G Q+  + A+++   ++ I
Sbjct: 121 ELRPTPELSFAVRKLGAAGGIVITASHNPPEYNGFKVYGSDGCQLPPEPANRLVKFVNEI 180

Query: 184 TDYQQIKQIPFEEALASGSASYIDESIEEAYKKEVLGLTINDTNID---KSVRVVYTPLN 240
            D   I      E   +G+   I E ++ AY + +  + +N   ++    K VR+V+TPL+
Sbjct: 181 EDELVIPVGDERELKENGTLEMIGEEVDVAYHEALKTIIVNPELLEASAKDVRIVFTPLH 240

Query: 241 GVGNLPVREVLRRRGFENVYVVPEQEMPDPDFTTVGYPNPEVPKAFAYSESLGKSVDADI 300
           G  NLPVR VL   GFENV VV EQE+PDP F+TV  PNPE   AFA +   GK  +AD+
Sbjct: 241 GTANLPVRRVLEAVGFENVTVVKEQELPDPQFSTVKAPNPEEHAAFALAIEYGKKTEADV 300

Query: 301 LLATDPDCDRVALEVKDSKGEYIFLNGNKIGALLSYYIFSQRCALGNLPHHPVLVKSIVT 360
           L+ATDPD DRV + V++  GEYI L GN+ G L+ +Y+ SQ+   G LP + + +K+IVT
Sbjct: 301 LIATDPDADRVGVAVQNQAGEYIVLTGNQTGGLMLHYLLSQKKEKGQLPVNGIALKTIVT 360

Query: 361 GDLSKVIADKYNIETVETLTGFKNICGKANEYDISKDKTYLFGYEESIGFCYGTFVRDKD 420
            +  + IA+ + I  V+TLTGFK I  K  EY+ S +  +LFGYEES G+  G FVRDKD
Sbjct: 361 SEFGRAIAEDFGIPMVDTLTGFKFIGEKIKEYEQSGEHQFLFGYEESYGYLIGDFVRDKD 420

Query: 421 AVSASMMVVEMTAYYKERGQTLLDVLQTIYDKFGYYNERQFSLELEGAEGQERISRIMED 480
           AV A ++  EMTAYYK RG TL D L  ++D++GYY E   S+ L+G  G E+I  ++
Sbjct: 421 AVQACLLAAEMTAYYKSRGMTLYDGLLELFDRYGYYREGLTSITLKGKVGVEKIQHVLSQ 480

Query: 481 FRQDPILQVGEMTLENSIDFKDGYK----------DFPKQNCLKYYFNEGSWYALRPSG 529
           FRQ P  QV +  +    D++   K             P  N LKY   +GSW+ LRPSG
Sbjct: 481 FRQSPPKQVNDQQVVVIEDYQTKEKVSVKERTVEAITLPTSNVLKYMLEDGSWFCLRPSG 540

Query: 530 TEPKIKCY 537
           TEPK+K Y
Sbjct: 541 TEPKLKIY 548
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 265> which encodes the amino acid sequence <SEQ ID 266>. Analysis of this protein sequence reveals the following:

```
Possible site: 35

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.5497(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 470/564 (83%), Positives = 517/564 (91%)

Query: 1   MSHMNYKEIYQEWLENDSLGKDIKSDLEAIKGDESEIQDRFYKTLEFGTAGLRGKLGAGT 60
           MS+M Y E+YQEWL N+ L  DIK+DL AIK +E+EIQDRFYKTLEFGTAGLRGKLGAGT
Sbjct: 1   MSNMTYNEVYQEWLHNNDLSDDIKADLAAIKDNEAEIQDRFYKTLEFGTAGLRGKLGAGT 60

Query: 61  NRMNTYMVGKAAQALANTIIDHGPEAIARGIAVSYDVRYQSKEFAELTCSIMAANGIKSY 120
           NRMNTYMVGKAAQALANTIIDHGPEA+ +GIAVSYDVRYQS+ FAELTCSIMAANGIK+Y
Sbjct: 61  NRMNTYMVGKAAQALANTIIDHGPEAVKKGIAVSYDVRYQSRTFAELTCSIMAANGIKAY 120

Query: 121 IYKGIRPTPMCSYAIRALGCVSGVMITASHNPQAYNGYKAYWKEGSQILDDIADQIANHM 180
           +YKGIRPTPMCSYAIRALGC+SGVMITASHNPQAYNGYKAYW+EGSQILDDIADQIA HM
Sbjct: 121 LYKGIRPTPMCSYAIRALGCISGVMITASHNPQAYNGYKAYWQEGSQILDDIADQIAQHM 180

Query: 181 DAITDYQQIKQIPFEEALASGSASYIDESIEEAYKKEVLGLTINDTNIDKSVRVVYTPLN 240
            A+T YQ+IKQ+PFE+AL SG  +YIDESIEEAYKKEVLGLTINDT+IDKSVRVVYTPLN
Sbjct: 181 AALTQYQEIKQMPFEKALDSGLVTYIDESIEEAYKKEVLGLTINDTDIDKSVRVVYTPLN 240

Query: 241 GVGNLPVREVLRRRGFENVYVVPEQEMPDPDFTTVGYPNPEVPKAFAYSESLGKSVDADI 300
           GVGNLPVREVLRRRGFENVYVVPEQEMPDPDFTTVGYPNPEVPK FAYSE LGK+VDADI
Sbjct: 241 GVGNLPVREVLRRRGFENVYVVPEQEMPDPDFTTVGYPNPEVPKTFAYSEKLGKAVDADI 300

Query: 301 LLATDPDCDRVALEVKDSKGEYIFLNGNKIGALLSYYIFSQRCALGNLPHHPVLVKSIVT 360
           L+ATDPDCDRVALEVK++ G+Y+FLNGNKIGALLSYYIFSQR  LGNLP +PVLVKSIVT
Sbjct: 301 LIATDPDCDRVALEVKNAVGDYVFLNGNKIGALLSYYIFSQRFDLGNLPANPVLVKSIVT 360

Query: 361 GDLSKVIADKYNIETVETLTGFKNICGKANEYDISKDKTYLFGYEESIGFCYGTFVRDKD 420
           GDLS+ IA  Y IETVETLTGFKNICGKANEYD++K K YLFGYEESIGFCYGTFVRDKD
Sbjct: 361 GDLSRAIASHYGIETVETLTGFKNICGKANEYDVTKQKNYLFGYEESIGFCYGTFVRDKD 420

Query: 421 AVSASMMVVEMTAYYKERGQTLLDVLQTIYDKFGYYNERQFSLELEGAEGQERISRIMED 480
           AVSASMM+VEM AYYK++GQ LLDVLQTIY  FGYYNERQ +LELEG EGQ+RI+RIMED
Sbjct: 421 AVSASMMIVEMAAYYKKKGQNLLDVLQTIYATFGYYNERQIALELEGIEGQKRIARIMED 480

Query: 481 FRQDPILQVGEMTLENSIDFKDGYKDFPKQNCLKYYFNEGSWYALRPSGTEPKIKCYLYT 540
           FRQ PI  V EM L+ +IDF DGY+DFPKQNCLK+Y ++GSWYALRPSGTEPKIK YLYT
Sbjct: 481 FRQTPIASVAEMALDKTIDFIDGYQDFPKQNCLKFYLDDGSWYALRPSGTEPKIKFYLYT 540

Query: 541 IGCTEADSLSKLNAIESACRAKMN 564
           IG T+ +S +KL+AIE+ACR K+N
Sbjct: 541 IGQTQENSATKLDAIEAACRTKIN 564
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 80

A DNA sequence (GBSx0080) was identified in *S.agalactiae* <SEQ ID 267> which encodes the amino acid sequence <SEQ ID 268>. This protein is predicted to be methylenetetrahydrofolate dehydrogenase (folD). Analysis of this protein sequence reveals the following:

```
Possible site: 48

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4672(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC44612 GB:U58210 tetrahydrofolate dehydrogenase/cyclohydrolase
          [Streptococcus thermophilus]
 Identities = 209/282 (74%), Positives = 248/282 (87%)
```

```
Query: 1   MTELIDGKALSQKMQAELGRKVERLKEQHGIIPGLAVILVGDNPASQVYVRNKERSALEA 60
           M  ++DGKAL+  MQ +L  KV RLKE+  I+PGL VI+VG+NPASQVYVRNKER+A +A
Sbjct: 1   MAIIMDGKALAVNMQEQLQEKVARLKEKEWIVPGLVVIMVGENPASQVYVRNKERAAKKA 60

Query: 61  GFKSETLRLSESISQEELIDIIHQYNEDKSIHGILVQLPLPQHINDKKIILAIDPKKDVD 120
           GF S+T+ LSESIS+EELI++I +YN++   HGILVQLPLP HIN+ +I+LAIDPKKDVD
Sbjct: 61  GFHSKTVNLSESISEEELIEVIEKYNQNPLFHGILVQLPLPNHINEMRILLAIDPKKDVD 120

Query: 121 GFHPMNTGHLWSGRPMMVPCTPAGIMEMFREYHVDLEGKHAVIIGRSNIVGKPMAQLLLD 180
           GFHPMNTG+LW+GRP MVPCTPAGIME+ REY+V+LEGK AVIIGRSNIVGKPMAQLLL+
Sbjct: 121 GFHPMNTGNLWNGRPQMVPCTPAGIMEILREYNVELEGKTAVIIGRSNIVGKPMAQLLLE 180

Query: 181 KNATVTLTHSRTRNLSEVTKEADILIVAIGQGHFVTKDFVKEGAVVIDVGMNRDENGKLI 240
           KNATVTLTHSRT +L++V  +AD+LIVAIG+  FVT++FVKEGAVVIDVG+NRDE GKL
Sbjct: 181 KNATVTLTHSRTPHLAKVCNKADVLIVAIGRAKFVTEEFVKEGAVVIDVGINRDEEGKLC 240

Query: 241 GDVVFEQVAEVASMITPVPGGVGPMTITMLLEQTYQAALRSV 282
           GDV F+QV E  SMITPVPGGVGPMTITML+EQTYQAALRS+
Sbjct: 241 GDVDFDQVKEKVSMITPVPGGVGPMTITMLMEQTYQAALRSL 282
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 269> which encodes the amino acid sequence <SEQ ID 270>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3368(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 230/281 (81%), Positives = 257/281 (90%)

Query: 1   MTELIDGKALSQKMQAELGRKVERLKEQHGIIPGLAVILVGDNPASQVYVRNKERSALEA 60
           MTELIDGKAL+QKMQ EL  KV  LK++ GI+PGLAVILVGD+PASQVYVRNKER+AL
Sbjct: 3   MTELIDGKALAQKMQQELAAKVNNLKQKKGIVPGLAVILVGDDPASQVYVRNKERAALTV 62

Query: 61  GFKSETLRLSESISQEELIDIIHQYNEDKSIHGILVQLPLPQHINDKKIILAIDPKKDVD 120
           GFKSET+RLSE I QEELI +I +YN D +IHGILVQLPLP HINDKKIILAIDPKKDVD
Sbjct: 63  GFKSETVRLSEFICQEELIAVIERYNADNTIHGILVQLPLPNHINDKKIILAIDPKKDVD 122

Query: 121 GFHPMNTGHLWSGRPMMVPCTPAGIMEMFREYHVDLEGKHAVIIGRSNIVGKPMAQLLLD 180
           GFHPMNTGHLWSGRP+MVPCTP+GIME+ REY+V+LEGKHAVIIGRSNIVGKPMAQLLLD
Sbjct: 123 GFHPMNTGHLWSGRPLMVPCTPSGIMELLREYNVNLEGKHAVIIGRSNIVGKPMAQLLLD 182

Query: 181 KNATVTLTHSRTRNLSEVTKEADILIVAIGQGHFVTKDFVKEGAVVIDVGMNRDENGKLI 240
           KNATVTLTHSRTR L EV + AD+LIVAIGQGHF+TK ++K+GA+VIDVGMNRD+NGKLI
Sbjct: 183 KNATVTLTHSRTRQLEEVCRCADVLIVAIGQGHFITKQYIKDGAIVIDVGMNRDDNGKLI 242

Query: 241 GDVVFEQVAEVASMITPVPGGVGPMTITMLLEQTYQAALRS 281
           GDV F++VAEVA+ ITPVPGGVGPMTI MLLEQTYQ+ALRS
Sbjct: 243 GDVAFDEVAEVAAKITPVPGGVGPMTIAMLLEQTYQSALRS 283
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 81

A DNA sequence (GBSx0081) was identified in *S.agalactiae* <SEQ ID 271> which encodes the amino acid sequence <SEQ ID 272>. Analysis of this protein sequence reveals the following:

```
Possible site: 39

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.24   Transmembrane   39 -  55 (  38 -  58)

----- Final Results -----
              bacterial membrane --- Certainty=0.2296(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9623> which encodes amino acid sequence <SEQ ID 9624> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC44613 GB:U58210 orf1091 [Streptococcus thermophilus]
 Identities = 149/277 (53%), Positives = 191/277 (68%)

Query: 1   MIVGEQEARALIKPRPKSSHKGDYGSVLLIGGFYPYGGAIIMAALACVKTGAGLVTVATQ 60
           M V +   R +I+PR + SHKG YG VLL+GG YPYGGAIIMAA+ACV +GAGLVTVAT
Sbjct: 1   MKVDDDLVRQVIRPRLRGSHKGSYGRVLLVGGLYPYGGAIIMAAIACVNSGAGLVTVATD 60

Query: 61  SCNIPSLHSQLPEVMAFDSDDYKWLEKSIVQSDVIVIGPGLGVSESSRKILNQTMEKIQS 120
             NI +LH+ LPE MAFD  + +     +  +DVI+IG GLG  E++   L   +  I+S
Sbjct: 61  RENIIALHAHLPEAMAFDLRETERFLDKLRAADVILIGSGLGEEETADWALELVLANIRS 120

Query: 121 HQSVILDGSALTLLSEGAFPQTKAKNLVLTPHQKEWERLSGIAVSQQTKENTQTALKSFP 180
           +Q++++DGSAL LL++          +L+LTPHQKEWERLSG+A+S+Q+  NTQ AL+ F
Sbjct: 121 NQNLVVDGSALNLLAKKNQSSLPKCHLILTPHQKEWERLSGLAISEQSVSNTQRALEEFQ 180

Query: 181 KGTILVAKSSHTRIFQDLDEKEIIVGGPYQATGGMGDTLCGMIAGMLAQFKEASPLDKVS 240
            GTILVAKS  T ++Q  +   + VGGPYQATGGMGDTL GM+AG LAQF         V
Sbjct: 181 SGTILVAKSHKTAVYQGAEVTHLEVGGPYQATGGMGDTLAGMVAGFLAQFASTDSYKAVI 240

Query: 241 VGVYLHSAIAQGLSKEAYVVLPTTISDEIPKEMARLS 277
           V  +LHSAIA  +++ AYVVLPT IS  IP  M +LS
Sbjct: 241 VATWLHSAIADNIAENAYVVLPTRISKAIPSWMKKLS 277
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 272 (GBS413) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 79 (lane 2; MW 34.2kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 171 (lane 7; MW 59kDa).

GBS413-GST was purified as shown in Figure 218, lane 12.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 82

A DNA sequence (GBSx0082) was identified in *S.agalactiae* <SEQ ID 273> which encodes the amino acid sequence <SEQ ID 274>. This protein is predicted to be Exonuclease VII large subunit (xseA). Analysis of this protein sequence reveals the following:

```
Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3172(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB14361 GB:Z99116 similar to exodeoxyribonuclease VII (large
          subunit) [Bacillus subtilis]
 Identities = 193/446 (43%), Positives = 283/446 (63%), Gaps = 10/446 (2%)

Query: 4   YLSVSTLTKYLKLKFDKDPYLERVYLTGQVSNFR-RRPNHQYFSLKDDKSVIQATMWSGH 62
           Y++VS LTKY+K KFD DP+LE +++ G++SN +     H YF+LK+ K  +Q+ M++
Sbjct: 6   YVTVSALTKYIKRKFDVDPHLENIWIKGELSNVKIHTRGHIYFTLKERKGRMQSVMFARQ 65

Query: 63  FKKLGFELEEGMKVNVVGRVQLYEPSGSYSIIVEKAEPDGIGALAIQFEQLKKKLSQAGY 122
            ++L F+ E GMKV V G + +YEPSG+Y +  ++ +PDG+GAL + +E+LKKKL+  G
Sbjct: 66  SERLPFKPENGMKVLVRGGISVYEPSGNYQLYAKEMQPDGVGALYLAYEELKKKLAGEGL 125

Query: 123 FDDRHKQLIPQFVRKIGVVTSPSGAVIRDIITTVSRRFPGVEILLFPTKVQGEGAAQEIA 182
           FDDR+K+ IP F   IGVVTSP+GA +RD+ITT+ RR+P V++++ P  VQGE A++ I
Sbjct: 126 FDDRYKKQIPAFPATIGVVTSPTGAAVRDVITTLKRRYPLVKVIVLPALVQGENASRSIV 185

Query: 183 QTIALANEKKDLDLLIVGRGGGSIEDLWAFNEECVVEAIFESRLPVISSVGHETDTTLAD 242
             I  ANEK+  D+LIVGRGGGSIE+LWAFNEE V  AIF S +P+IS+VGHETD T++D
Sbjct: 186 TRIEEANEKEICDVLIVGRGGGSIEELWAFNEEIVARAIFASNIPIISAVGHETDFTISD 245

Query: 243 FVADRRAATPTAAAELATPVTKIDILSWITERENRMYQSSLRLIRTKEERLQKSKQSVIF 302
           FVAD RAATPT AAE+A P T  D++      E RM ++  + +  ++ R+Q  + S  F
Sbjct: 246 FVADIRAATPTGAAEIAVPHT-TDLIERTKTAEVRMTRAMQQHLGQEKGRIQTLQSSYAF 304

Query: 303 RQPERLYDGFLQKLD----NLNQQLTYSMRDKLQTVRQKQGLLHQKLQGIDLKQRIHIYQ 358
           R P+RLY    Q+ D        QLT  +  K + + ++   L        LKQ    YQ
Sbjct: 305 RFPKRLYAQKEQQFDLAYQQFQAQLTALLDRKSRQLERETYRLEALHPHEQLKQARTRYQ 364

Query: 359 ERVVQSRRLLSSTMTSQYDSKLARFEKAQDALISLDSSRIVARGYAIIEKNHTLVSTTNG 418
           E+  Q R+     M  Q     ++F+     L +L   +++ RGY++  K   L+ + +
Sbjct: 365 EQTNQLRK----NMNIQMKQLHSQFQTVLGKLNALSPLQVMERGYSLAYKEDKLIKSVSQ 420

Query: 419 INEGDHLQVKMQDGLLEVEVKDVRQE 444
           I E D L++K++DG+L  EV + R E
Sbjct: 421 IEEQDRLEIKLKDGVLTCEVLEKRGE 446
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 275> which encodes the amino acid sequence <SEQ ID 276>. Analysis of this protein sequence reveals the following:

```
Possible site: 61

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3275(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 321/446 (71%), Positives = 386/446 (85%)

Query: 1   MSDYLSVSTLTKYLKLKFDKDPYLERVYLTGQVSNFRRRPNHQYFSLKDDKSVIQATMWS 60
           M+DYL+V+ LTKYLKLKFD+DPYLERVYLTGQVSNFR+RP HQYFSLKD+ +VIQATMW+
Sbjct: 6   MADYLTVTHLTKYLKLKFDRDPYLERVYLTGQVSNFRKRPTHQYFSLKDESAVIQATMWA 65

Query: 61  GHFKKLGFELEEGMKVNVVGRVQLYEPSGSYSIIVEKAEPDGIGALAIQFEQLKKKLSQA 120
           G +KKLGF+LEEGMK+NV+GRVQLYEPSGSYSI++EKAEPDGIGALA+QFEQLKKKL+
Sbjct: 66  GVYKKLGFDLEEGMKINVIGRVQLYEPSGSYSIVIEKAEPDGIGALALQFEQLKKKLTAE 125

Query: 121 GYFDDRHKQLIPQFVRKIGVVTSPSGAVIRDIITTVSRRFPGVEILLFPTKVQGEGAAQE 180
           GYF+ +HKQ +PQFV KIGV+TSPSGAVIRDIITTVSRRFPGVEILLFPTKVQG+GAAQE
Sbjct: 126 GYFEQKHKQPLPQFVSKIGVITSPSGAVIRDIITTVSRRFPGVEILLFPTKVQGDGAAQE 185
```

```
Query: 181 IAQTIALANEKKDLDLLIVGRGGGSIEDLWAFNEECVVEAIFESRLPVISSVGHETDTTL 240
           +   I  AN+++DLDLLIVGRGGGSIEDLWAFNEE VV+AIFES+LPVISSVGHETDTTL
Sbjct: 186 VVANIRRANQREDLDLLIVGRGGGSIEDLWAFNEEIVVQAIFESQLPVISSVGHETDTTL 245

Query: 241 ADFVADRRAATPTAAAELATPVTKIDILSWITERENRMYQSSLRLIRTKEERLQKSKQSV 300
           ADFVADRRAATPTAAAELATP+TK D++SWI ER+NR YQ+ LR I+ ++E + K  QSV
Sbjct: 246 ADFVADRRAATPTAAAELATPITKTDLMSWIVERQNRSYQACLRRIKQRQEWVDKLSQSV 305

Query: 301 IFRQPERLYDGFLQKLDNLNQQLTYSMRDKLQTVRQKQGLLHQKLQGIDLKQRIHIYQER 360
           IFRQPERLYD +LQK+D L+  L  +M+D+L + ++ +  L   L    L+ +I  YQ+R
Sbjct: 306 IFRQPERLYDAYLQKIDRLSMTLMNTMKDRLSSAKENKVQLDHALANSQLQTKIERYQDR 365

Query: 361 VVQSRRLLSSTMTSQYDSKLARFEKAQDALISLDSSRIVARGYAIIEKNHTLVSTTNGIN 420
           V  ++RLL + M SQYDS+LARFEKAQDAL+SLD+SRI+ARGYA+IEKN  LV++ + I
Sbjct: 366 VATAKRLLMANMASQYDSQLARFEKAQDALLSLDASRIIARGYAMIEKNQALVASVSQIT 425

Query: 421 EGDHLQVKMQDGLLEVEVKDVRQENI 446
           +GD L +KM+DG L+VEVKDV+ ENI
Sbjct: 426 KGDQLTIKMRDGQLDVEVKDVKNENI 451
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 83

A DNA sequence (GBSx0083) was identified in *S.agalactiae* <SEQ ID 277> which encodes the amino acid sequence <SEQ ID 278>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2913(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAG07429 GB:AE004821 exodeoxyribonuclease VII small subunit
          [Pseudomonas aeruginosa]
 Identities = 26/66 (39%), Positives = 51/66 (76%), Gaps = 2/66 (3%)

Query: 1  MSDKKT--FEENLQELETIVSRLETGDVALEDAIAEFQKGMLISKELQRTLKEAEETLVK 58
          M+ KKT  FE++L EL+T+V RLE+G+++LE+++  F++G+ +++E Q +L +AE+ +
Sbjct: 1  MARKKTLDFEQSLTELQTLVERLESGELSLEESLGAFEQGIRLTRECQTSLSQAEQKVQI 60

Query: 59 VMQADG 64
          +++ DG
Sbjct: 61 LLERDG 66
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 279> which encodes the amino acid sequence <SEQ ID 280>. Analysis of this protein sequence reveals the following:

```
Possible site: 51

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2796(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 55/70 (78%), Positives = 65/70 (92%)

Query: 1  MSDKKTFEENLQELETIVSRLETGDVALEDAIAEFQKGMLISKELQRTLKEAEETLVKVM 60
          MS  KTFEENLQ+LETIV++LE GDV LE+AI+EFQKGML+SKELQ+TL+ AE+TLVKVM
Sbjct: 1  MSKTKTFEENLQDLETIVNKLENGDVPLEEAISEFQKGMLLSKELQKTLQAAEKTLVKVM 60

Query: 61 QADGTEVEMD 70
          QADGTEV+MD
Sbjct: 61 QADGTEVDMD 70
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 84

A DNA sequence (GBSx0084) was identified in *S.agalactiae* <SEQ ID 281> which encodes the amino acid sequence <SEQ ID 282>. Analysis of this protein sequence reveals the following:

```
Possible site: 58

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2614(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA25265 GB:AB003187 farnesyl diphosphate synthase [Micrococcus
          luteus]
 Identities = 126/258 (48%), Positives = 175/258 (66%), Gaps = 2/258 (0%)

Query: 27  LIKAILYSVDGGGKRIRPRILLEILEGFGVELIDGHYDVAAALEMIHTGSLIHDDLPAMD 86
          L +AI YS+  GGKRIRP ++L  L+  G    DG      ALEMIHT SLIHDDLPAMD
Sbjct: 31  LHEAINYSLSAGGKRIRPLLVLTTLDSLGGNAHDG-LPFGIALEMIHTYSLIHDDLPAMD 89

Query: 87  NDDFRRGRLTNHKKFDEATAVLAGDSLFLDPFDLVVKAGFKADVTVRLIELLSMSAGSFG 146
          NDD+RRG+LTNHK+FDEATA+LAGD+L  D F  ++     A++ + LI LLS ++GS G
Sbjct: 90  NDDYRRGKLTNHKRFDEATAILAGDALLTDAFQCILNTQLNAEIKLSLINLLSTASGSNG 149

Query: 147 MVGGQMLDMKGENKVLSIDDLSLIHINKTGRLLAYPFVAAGILAEKSEEVKGKLHQAGLL 206
          MV GQMLDM+GE+K L++++L  IHI+KTG L+    V+AGI+   ++    +L+  G
Sbjct: 150 MVYGQMLDMQGEHKTLTLNELERIHIHKTGELIRAAIVSAGIIMNFNDAQIEQLNIIGKN 209

Query: 207 IGHAFQVRDDILDVTASFEELGKTPNKDIVAEKTTYPNLLGLDKSQEILDDTLKKAQAIF 266
          +G  FQ++DDILDV  SFE +GKT   D+  +K+TY +LLGL+ S+++L+D L +
Sbjct: 210 VGLMFQIKDDILDVEGSFENIGKTVGSDLNNDKSTYVSLLGLEASKQLLNDKLTETYDAL 269

Query: 267 QNLEKKANFNARKIIDII 284
          + L+   N N + +I  I
Sbjct: 270 KTLQ-PINDNLKTLITYI 286
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 283> which encodes the amino acid sequence <SEQ ID 284>. Analysis of this protein sequence reveals the following:

```
Possible site: 38

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3887(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 192/289 (66%), Positives = 237/289 (81%)

Query: 2   MVTIEKIDEAIHRYYKQTHSVVSPDLIKAILYSVDGGGKRIRPRILLEILEGFGVELIDG 61
           M  + +IDEAI RYYK T + VS +LI AILYSVD GGKRIRP ILLE++EGFGV L +
Sbjct: 1   MDKLARIDEAIRRYYKTTSNGVSEELIDAILYSVDSGGKRIRPLILLEMIEGFGVSLQNA 60

Query: 62  HYDVAAALEMIHTGSLIHDDLPAMDNDDFRRGRLTNHKKFDEATAVLAGDSLFLDPFDLV 121
           H+D+AAALEMIHTGSLIHDDLPAMDNDD+RRGRLTNHK+F EATA+LAGDSLFLDPF L+
Sbjct: 61  HFDLAAALEMIHTGSLIHDDLPAMDNDDYRRGRLTNHKQFGEATAILAGDSLFLDPFGLI 120

Query: 122 VKAGFKADVTVRLIELLSMSAGSFGMVGGQMLDMKGENKVLSIDDLSLIHINKTGRLLAY 181
            +A   ++V V LI+ LS+++G+FGMVGGQMLDMKGEN+ LS+  LSLIH+NKTG+LLA+
Sbjct: 121 AQAELNSEVKVALIQELSLASGTFGMVGGQMLDMKGENQALSLPQLSLIHLNKTGKLLAF 180

Query: 182 PFVAAGILAEKSEEVKGKLHQAGLLIGHAFQVRDDILDVTASFEELGKTPNKDIVAEKTT 241
           PF AA ++ E++  V+ +L QAG+LIGHAFQ+RDDILDVTASFE+LGKTP KD+ AEK T
Sbjct: 181 PFKAAALITEQAMTVRQQLEQAGMLIGHAFQIRDDILDVTASFEDLGKTPKKDLFAEKAT 240

Query: 242 YPNLLGLDKSQEILDDTLKKAQAIFQNLEKKANFNARKIIDIIEGLRLN 290
           YP+LLGL+ S ++L ++L +A  IFQ LE    F  + I  +IEGLRLN
Sbjct: 241 YPSLLGLEASYQLLTESLDQALTIFQTLESDVGFKPQIITKLIEGLRLN 289
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 85

A DNA sequence (GBSx0085) was identified in *S.agalactiae* <SEQ ID 285> which encodes the amino acid sequence <SEQ ID 286>. This protein is predicted to be hemolysin-like protein (tly). Analysis of this protein sequence reveals the following:

```
Possible site: 37

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.75   Transmembrane  152 - 168 ( 151 - 168)

----- Final Results -----
               bacterial membrane --- Certainty=0.1298(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB06497 GB:AP001516 hemolysin-like protein [Bacillus halodurans]
 Identities = 162/270 (60%), Positives = 202/270 (74%), Gaps = 3/270 (1%)

Query: 3   KERVDVLAYKQGLFDTREQAKRGVMAGMVINVINGERYDKPGEKVADDTELKLKGEKLKY 62
           KERVDVL  ++GL +TRE+AKR +MAG+V +    ER DKPG KV  DT L +KGE L Y
Sbjct: 4   KERVDVLLVERGLMETREKAKRSIMAGLVFS--GHERVDKPGLKVDRDTPLSVKGEVLPY 61

Query: 63  VSRGGLKLEKALQVFEISVADKLTIDIGASTGGFTDVMLQSGARLVYAVDVGTNQLVWKL 122
           VSRGGLKLEKA++ F++ + D++ +DIGASTGGFTD  LQ+GA  VYAVDVG NQL WKL
Sbjct: 62  VSRGGLKLEKAIRAFDLHLTDRVVLDIGASTGGFTDCALQNGATFVYAVDVGYNQLAWKL 121

Query: 123 RQDHRVRSMEQYNFRYAQKEDFKEGLPEFASIDVSFISLNLILPALKEILVDGGQVVALI 182
           RQD RV  ME+ NFRY + E  + GLP  A+IDVSFISL LILP LK +L++   VVAL+
Sbjct: 122 RQDERVVVMERTNFRYLKPEVLERGLPNMATIDVSFISLKLILPVLKTMLLENSDVVALV 181

Query: 183 KPQFEAGREQIGKNGIVKDKLVHEKVLTTVTNFTKDYGYTVKHLDFSPIQGGHGNIEFLM 242
           KPQFEAGRE++GK GIV+DK VH+KVL+T+  F    GY V  LDFSPI GG GNIEFL+
Sbjct: 182 KPQFEAGREEVGKKGIVRDKSVHQKVLSTIVEFALKEGYAVGGLDFSPITGGEGNIEFLL 241

Query: 243 HLQKCQDPQNLV-LDQIQDVIEKAHKEFKK 271
```

```
            HL   +D ++ +  + I+D +E+AH E KK
Sbjct: 242 HLMWRKDKESFISQEMIRDTVERAHLELKK 271
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 287> which encodes the amino acid sequence <SEQ ID 288>. Analysis of this protein sequence reveals the following:

```
      Possible site: 37

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.92   Transmembrane  150 - 166 ( 149 - 168)

----- Final Results -----
               bacterial membrane --- Certainty=0.2168(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB06497 GB:AP001516 hemolysin-like protein [Bacillus halodurans]
 Identities = 156/270 (57%), Positives = 196/270 (71%), Gaps = 3/270 (1%)

Query: 3   KERVDVLAYKQGLFETREQAKRGVMAGLVVSVINGQRYDKPGDKIDDGTELKLKGEKLKY 62
           KERVDVL  ++GL ETRE+AKR +MAGLV S    +R DKPG K+D  T L +KGE L Y
Sbjct: 4   KERVDVLLVERGLMETREKAKRSIMAGLVFS--GHERVDKPGLKVDRDTPLSVKGEVLPY 61

Query: 63  VSRGGLKLEKGLHVFGVSVANQIGIDIGASTGGFTDVMLQDGAKLVYAVDVGTNQLVWKL 122
           VSRGGLKLEK +  F + + +++ +DIGASTGGFTD  LQ+GA  VYAVDVG NQL WKL
Sbjct: 62  VSRGGLKLEKAIRAFDLHLTDRVVLDIGASTGGFTDCALQNGATFVYAVDVGYNQLAWKL 121

Query: 123 RQDPRVRSMEQYNFRYAQPEDFNEGQPVFASIDVSFISLSLILPALHNVLSDQGQVIALI 182
           RQD RV  ME+ NFRY +PE    G P  A+IDVSFISL LILP L  +L +   V+AL+
Sbjct: 122 RQDERVVVMERTNFRYLKPEVLERGLPNMATIDVSFISLKLILPVLKTMLLENSDVVALV 181

Query: 183 KPQFEAGREQIGKKGIVKDKQIHEKVIQKVMDFASGYGFTVKGLDFSPIQGGHGNIEFLA 242
           KPQFEAGRE++GKKGIV+DK +H+KV+  +++FA   G+ V GLDFSPI GG GNIEFL
Sbjct: 182 KPQFEAGREEVGKKGIVRDKSVHQKVLSTIVEFALKEGYAVGGLDFSPITGGEGNIEFLL 241

Query: 243 HLAKSQTPET-LAPHLIQKVVAKAHKEFEK 271
           HL   +  E+ ++  +I+  V +AH E +K
Sbjct: 242 HLMWRKDKESFISQEMIRDTVERAHLELKK 271
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 214/275 (77%), Positives = 238/275 (85%)

Query: 1   MAKERVDVLAYKQGLFDTREQAKRGVMAGMVINVINGERYDKPGEKVADDTELKLKGEKL 60
           M KERVDVLAYKQGLF+TREQAKRGVMAG+V++VING+RYDKPG+K+ D TELKLKGEKL
Sbjct: 1   MPKERVDVLAYKQGLFETREQAKRGVMAGLVVSVINGQRYDKPGDKIDDGTELKLKGEKL 60

Query: 61  KYVSRGGLKLEKALQVFEISVADKLTIDIGASTGGFTDVMLQSGARLVYAVDVGTNQLVW 120
           KYVSRGGLKLEK L VF +SVA+++ IDIGASTGGFTDVMLQ GA+LVYAVDVGTNQLVW
Sbjct: 61  KYVSRGGLKLEKGLHVFGVSVANQIGIDIGASTGGFTDVMLQDGAKLVYAVDVGTNQLVW 120

Query: 121 KLRQDHRVRSMEQYNFRYAQKEDFKEGLPEFASIDVSFISLNLILPALKEILVDGGQVVA 180
           KLRQD RVRSMEQYNFRYAQ EDF EG P FASIDVSFISL+LILPAL  +L D GQV+A
Sbjct: 121 KLRQDPRVRSMEQYNFRYAQPEDFNEGQPVFASIDVSFISLSLILPALHNVLSDQGQVIA 180

Query: 181 LIKPQFEAGREQIGKNGIVKDKLVHEKVLTTVINFTKDYGYTVKHLDFSPIQGGHGNIEF 240
           LIKPQFEAGREQIGK GIVKDK +HEKV+  V +F   YG+TVK LDFSPIQGGHGNIEF
Sbjct: 181 LIKPQFEAGREQIGKKGIVKDKQIHEKVIQKVMDFASGYGFTVKGLDFSPIQGGHGNIEF 240

Query: 241 LMHLQKCQDPQNLVLDQIQDVIEKAHKEFKKNEEE 275
           L HL K Q P+ L    IQ V+ KAHKEF+K+E+E
Sbjct: 241 LAHLAKSQTPETLAPHLIQKVVAKAHKEFEKHEKE 275
```

SEQ ID 286 (GBS310) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 57 (lane 3; MW 34kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 61 (lane 4; MW 58.8kDa).

The GBS310-GST fusion product was purified (Figure 210, lane 10) and used to immunise mice. The resulting antiserum was used for FACS (Figure 282), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 86

A DNA sequence (GBSx0086) was identified in *S.agalactiae* <SEQ ID 289> which encodes the amino acid sequence <SEQ ID 290>. Analysis of this protein sequence reveals the following:

```
Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1966(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA09426 GB:AJ010954 arginine repressor [Bacillus
           stearothermophilus]
 Identities = 49/153 (32%), Positives = 84/153 (54%), Gaps = 4/153 (2%)

Query: 1   MKKSERLNLIKQIVLNHAVETQHELLRRLEAYGVTLTQATISRDMNEIGIIKVPSAKGRY 60
           M K +R   I++I++NH +ETQ EL+  L+  G  +TQAT+SRD+ E+ ++KVP A GRY
Sbjct: 1   MNKGQRHIKIREIIMNHEIETQDELVDMLKKAGFNVTQATVSRDIKELQLVKVPMANGRY 60

Query: 61  IYGLSNENDPIFTTAVAKPIKTSILSISDKLLGLEQFININVIPGNSQLIKTFIMSHCQE 120
            Y L   +D  F     + +K +++     KL G    + +  +PGN+  I   + +
Sbjct: 61  KYSL--PSDQRFNP--TQKLKRALMDAFVKLDGSGNLLVLKTLPGNAHAIGVLLDNLDWN 116

Query: 121 HIFSLTADDNSLLLIAKSEADADHIRQSMIAML 153
            I      D++ L+I ++  DA+ +   ++ ML
Sbjct: 117 EIVGTICGDDTCLIICRTAEDAEKVSGQLLGML 149
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 291> which encodes the amino acid sequence <SEQ ID 292>. Analysis of this protein sequence reveals the following:

```
Possible site: 50

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1717(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 87/154 (56%), Positives = 118/154 (76%), Gaps = 1/154 (0%)

Query: 1   MKKSERLNLIKQIVLNHAVETQHELLRRLEAYGVTLTQATISRDMNEIGIIKVPSAKGRY 60
           MKKSERL LIK++VL H +ETQH+LLR L  +G+ LTQATISRDMNEIGI+K+PS  GRY
Sbjct: 12  MKKSERLELIKKMVLTHPIETQHDLLRLLAEHGLELTQATISRDMNEIGIVKIPSGSGRY 71
```

```
Query: 61  IYGLSNENDPIFTTAVAKPIKTSILSISDKLLGLEQFININVIPGNSQLIKTFIMSHCQE 120
           IYGLS ++         + IK++IL++SDK  GLEQ + + V+PGNS+LIK ++++   +
Sbjct: 72  IYGLSQDSGKKIVQG-PRSIKSTILAVSDKTKGLEQHLYLKVVPGNSKLIKRYLLADFSK 130

Query: 121 HIFSLTADDNSLLLIAKSEADADHIRQSMIAMLE 154
            IFSL ADD+SLLLIAKS ++AD IRQ ++  ++
Sbjct: 131 AIFSLIADDDSLLLIAKSPSEADMIRQEILLWMQ 164
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 87

A DNA sequence (GBSx0088) was identified in *S.agalactiae* <SEQ ID 293> which encodes the amino acid sequence <SEQ ID 294>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3339(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 88

A DNA sequence (GBSx0089) was identified in *S.agalactiae* <SEQ ID 295> which encodes the amino acid sequence <SEQ ID 296>. This protein is predicted to be DNA repair protein recn (recN). Analysis of this protein sequence reveals the following:

```
Possible site: 50

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1651(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB14355 GB:Z99116 recN [Bacillus subtilis]
 Identities = 244/567 (43%), Positives = 366/567 (64%), Gaps = 18/567 (3%)

Query: 1   MLLEISIKNFAIIEEISLNFETGMTVLTGETGAGKSIIIDAMNMMLGSRASVEVIRHGAN 60
           ML E+SIKNFAIIEE++++FE G+TVLTGETGAGKSIIIDA+++++G R S E +R+G
Sbjct: 1   MLAELSIKNFAIIEELTVSFERGLTVLTGETGAGKSIIIDAISLLVGGRGSSEFVRYGEA 60

Query: 61  KAEIEGFFSVEKNQSLVQLLEENGIELADELII-RREIFQNGRSVSRINGQMVNLSTLKA 119
           KAE+EG F +E    ++ +  E GI+++DE+I+ RR+I  +G+SV R+NG++V +++L+
Sbjct: 61  KAELEGLFLLESGHPVLGVCAEQGIDVSDEMIVMRRDISTSGKSVCRVNGKLVTIASLRE 120

Query: 120 VGHYLVDIYGQHDQEELMKPNMHILMLDEFGNTEFNVIKERYQSLFDAYRQLRKRVLDKQ 179
           +G  L+DI+GQHD + LM+   H+ +LD+F   E     + YQ  +  Y +L K++
```

```
Sbjct: 121 IGRLLLDIHGQHDNQLLMEDENHLQLLDKFAGAEVESALKTYQEGYQRYVKLLKKLKQLS 180

Query: 180 KNEQENKSRIEMLEFQIAEIESVALKSDEDQTLLKQRDKLMNHKNIADTLTNAYLMLDNE 239
           ++EQE    +++++FQ+ EIES  L+ +ED+ L ++R ++ N + I ++L NAY  L +E
Sbjct: 181 ESEQEMAHCLDLIQFQLEEIESAKLELNEDEQLQEERQQISNFEKIYESLQNAYNALRSE 240

Query: 240 EFSSLSNVRSAMNDLMALEEFDREYKDLSTNLSEAYYVIEEVTKRLGDVIDDLDFDAGLL 299
           +   L  V  A   L  + + +   K +S ++S +YY++E+ T ++ +++D+L+FD   L
Sbjct: 241 Q-GGLDWVGMASAQLEDISDINEPLKKMSESVSNSYYLLEDATFQMRNMLDELEFDPERL 299

Query: 300 QEIENRLDVINTITRKYGGDVNDVLDYFDNITKEYSLLTGSEESSDALEKELKILEHDLI 359
             IE RL+ I  + RKYG  V D+L+Y   I +E   +   +    +L+KEL  +  D+
Sbjct: 300 NYIETRLNEIKQLKRKYGATVEDILEYASKIEEEIDQIENRDSHLQSLKKELDSVGKDVA 359

Query: 360 ESANQLSLERHKLAKQLENEIKQELTELYMEKADFQVQFTKG----------------KF 403
             A  +S  R   AK+L +EI +EL  LYMEK+ F  +F                    +
Sbjct: 360 VEAANVSQIRKTWAKKLADEIHRELKSLYMEKSTFDTEFKVRTASRNEEAPLVNGQPVQL 419

Query: 404 NKEGNEIVEFYISTNPGEGFKPLVKVASGGELSRLMLAIKSAFSRKEDKTSIVFDEVDTG 463
            ++G ++V+F ISTN GE  K L KVASGGELSR+MLAIKS FS ++D TSI+FDEVDTG
Sbjct: 420 TEQGIDLVKFLISTNTGEPLKSLSKVASGGELSRVMLAIKSIFSSQQDVTSIIFDEVDTG 479

Query: 464 VSGRVAQAIAQKIHKIGSHGQVLAISHLAQVIAIADYQYFIEKISSDSSTVSTVRLLSYE 523
           VSGRVAQAIA+KIHK+    QVL I+HL QV A+AD   +I K   D  T + V+ LS +
Sbjct: 480 VSGRVAQAIAEKIHKVSIGSQVLCITHLPQVAAMADTHLYIAKELKDGRTTTRVKPLSKQ 539

Query: 524 ERVEEIAKMLAGNNVTDTARTQAKELL 550
           E+V EI + +AG  VTD  +  AKELL
Sbjct: 540 EKVAEIERSIAGVEVTDLTKRHAKELL 566
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 297> which encodes the amino acid sequence <SEQ ID 298>. Analysis of this protein sequence reveals the following:

```
Possible site: 51

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1215(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 403/550 (73%), Positives = 472/550 (85%)

Query: 1   MLLEISIKNFAIIEEISLNFETGMTVLTGETGAGKSIIIDAMNMMLGSRASVEVIRHGAN 60
           MLLEISIKNFAII+EISLNFE GMTVLTGETGAGKSIIIDAMNMMLG+RAS EVIR GAN
Sbjct: 2   MLLEISIKNFAIIDEISLNFENGMTVLTGETGAGKSIIIDAMNMMLGARASTEVIRRGAN 61

Query: 61  KAEIEGFFSVEKNQSLVQLLEENGIELADELIIRREIFQNGRSVSRINGQMVNLSTLKAV 120
           KAEIEGFFSV+    LV  LE +GI + +ELIIRR+IF NGRSVSRINGQMVNL+TLK V
Sbjct: 62  KAEIEGFFSVDATPELVACLESSGIAMEEELIIRRDIFANGRSVSRINGQMVNLATLKQV 121

Query: 121 GHYLVDIYGQHDQEELMKPNMHILMLDEFGNTEFNVIKERYQSLFDAYRQLRKRVLDKQK 180
           G +LVDI+GQHDQEELM+P +H  +LD FG+  F  +KE YQ +FD Y+ LR++V+DKQK
Sbjct: 122 GQFLVDIHGQHDQEELMRPQLHQQILDAFGDKAFEQLKENYQLIFDRYKSLRRQVIDKQK 181

Query: 181 NEQENKSRIEMLEFQIAEIESVALKSDEDQTLLKQRDKLMNHKNIADTLTNAYLMLDNEE 240
           NE+E+K RI+ML FQIAEIE+ AL   ED  L ++RD+LMNHK IADTLTNAY+MLDN++
Sbjct: 182 NEKEHKDRIDMLAFQIAEIEAAALSRGEDDRLNQERDRLMNHKQIADTLTNAYVMLDNDD 241

Query: 241 FSSLSNVRSAMNDLMALEEFDREYKDLSTNLSEAYYVIEEVTKRLGDVIDDLDFDAGLLQ 300
           FSSLSN+RS+MNDL+++E+FD EYK +ST++SEAYY++EEV+K+L D ID LDFD G LQ
Sbjct: 242 FSSLSNIRSSMNDLLSIEQFDSEYKGMSTSISEAYYILEEVSKQLSDTIDQLDFDGGRLQ 301

Query: 301 EIENRLDVINTITRKYGGDVNDVLDYFDNITKEYSLLTGSEESSDALEKELKILEHDLIE 360
```

```
            EIE RLD++N++TRKYGG+VNDVLDY+DNI KEY LLTG + SS  LE ELK LE  L+
Sbjct: 302 EIEFRLDILNSLTRKYGGNVNDVLDYYDNIVKEYQLLTGDDLSSGDLEAELKSLEKQLVA 361

Query: 361 SANQLSLERHKLAKQLENEIKQELTELYMEKADFQVQFTKGKFNKEGNEIVEFYISTNPG 420
            +A++LS+ RH+LA+QLE EIK EL ELYMEKADF+V FT  KFN++GNE +EFYISTNPG
Sbjct: 362 AASELSVSRHQLAEQLEAEIKAELKELYMEKADFKVHFTTSKFNRDGNESLEFYISTNPG 421

Query: 421 EGFKPLVKVASGGELSRLMLAIKSAFSRKEDKTSIVFDEVDTGVSGRVAQAIAQKIHKIG 480
            EGFKPLVKVASGGELSRLMLAIK+A SRKEDKTSIVFDEVDTGVSGRVAQAIAQKI+KIG
Sbjct: 422 EGFKPLVKVASGGELSRLMLAIKAAISRKEDKTSIVFDEVDTGVSGRVAQAIAQKIYKIG 481

Query: 481 SHGQVLAISHLAQVIAIADYQYFIEKISSDSSTVSTVRLLSYEERVEEIAKMLAGNNVTD 540
             HGQVLAISHL QVIAIADYQYFI K S + STVS VRLL+ EERVEEIA M+AG ++T
Sbjct: 482 RHGQVLAISHLPQVIAIADYQYFISKESKEESTVSKVRLLTPEERVEEIASMIAGTDMTQ 541

Query: 541 TARTQAKELL 550
             A TQA+ELL
Sbjct: 542 AALTQARELL 551
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 89

A DNA sequence (GBSx0090) was identified in *S.agalactiae* <SEQ ID 299> which encodes the amino acid sequence <SEQ ID 300>. This protein is predicted to be degV protein. Analysis of this protein sequence reveals the following:

```
Possible site: 38

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.96   Transmembrane  246 - 262 ( 246 - 262)

----- Final Results -----
               bacterial membrane --- Certainty=0.1383(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB07346 GB:AP001519 unknown conserved protein [Bacillus halodurans]
 Identities = 93/277 (33%), Positives = 152/277 (54%), Gaps = 4/277 (1%)

Query: 1   MSKIKIVTDSSITIEPELIKELDITVVPLSVMIDGTLYSDNDLKAQGEFLNLMRGSKELP 60
           M+KI IVTDS+  + P+  KEL + VVPLSV+     Y +    +  +F   ++  ++LP
Sbjct: 1   MTKIAIVTDSTAYLGPKRAKELGVIVVPLSVVFGEEAYQEEVELSSADFYEKLKHEEKLP 60

Query: 61  KTSQPPVGVFAEIYEKLMNEGVEHIIAIHLTHTLSGTIE-ASRQGANIAGADVTVIDSTF 119
            TSQP VG+F E +E+L  EG E +I+IHL+  +SGT + A   G+ + G +V   DS
Sbjct: 61  TTSQPAVGLFVETFERLAKEGFEVVISIHLSSKISGTYQSALTAGSMVEGIEVIGYDSGI 120

Query: 120 TDQCQKFQVVEAAKLAKEGADLDTILARVEEVRQKSELFIGVSTLENLVKGGRIGRVTGL 179
           + + Q   V EAAKL KEGAD  TI+  ++EV++++     V  L +L +GGR+     +
Sbjct: 121 SCEPQANFVAEAAKLVKEGADPQTIIDHLDEVKKRTNALFVVHDLSHLHRGGRLNAAQLV 180

Query: 180 LSSLLNIKVIMELTNHELVPIVKGR-GLKTFSKWLDNFVESAQTRKIAEIGISYCGKADM 238
           + SLL IK I+   +  +VP+ K R   K +++  + F E A +    +  + +  + D
Sbjct: 181 VGSLLKIKPILHFEDGSIVPLEKVRTEKKAWARVKELFAEEASSASSVKATVIHANRLDG 240

Query: 239 ANNFREKL--AVLGAPISVLETGSIIQTHTGEDAFAV 273
             A    +++        +S+   G +I TH GE +  +
Sbjct: 241 AEKLADEIRSQFSHVDVSISHFGPVIGTHLGEGSIGL 277
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 301> which encodes the amino acid sequence <SEQ ID 302>. Analysis of this protein sequence reveals the following:

```
Possible site: 37

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.54   Transmembrane  180 - 196 ( 180 - 196)
   INTEGRAL    Likelihood = -0.16   Transmembrane   21 -  37 (  21 -  38)

----- Final Results -----
               bacterial membrane --- Certainty=0.1617(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 197/279 (70%), Positives = 226/279 (80%), Gaps = 1/279 (0%)

Query: 1   MSKIKIVTDSSITIEPELIKELDITVVPLSVMIDGTLYSDNDLKAQGEFLNLMRGSKELP 60
           M  IKIVTDSSITIEPELIK LDITVVPLSVMID  LYSDNDLK +G FL+LM+ SK LP
Sbjct: 5   MGTIKIVTDSSITIEPELIKALDITVVPLSVMIDSKLYSDNDLKEEGHFLSLMKASKSLP 64

Query: 61  KTSQPPVGVFAEIYEKLMNEGVEHIIAIHLTHTLSGTIEASRQGANIAGADVTVIDSTFT 120
           KTSQPPVG+FAE YE L+ +GV  I+AIHL+  LSGTIEASRQGA IA A VTV+DS FT
Sbjct: 65  KTSQPPVGLFAETYENLVKKGVTDIVAIHLSPALSGTIEASRQGAEIAEAPVTVLDSGFT 124

Query: 121 DQCQKFQVVEAAKLAKEGADLDTILARVEEVRQKSELFIGVSTLENLVKGGRIGRVTGLL 180
           DQ  KFQVVEAAK+AK GA L+ ILA V+ ++ K+EL+IGVSTLENLVKGGRIGRVTG+L
Sbjct: 125 DQAMKFQVVEAAKMAKAGASLNEILAAVQAIKSKTELYIGVSTLENLVKGGRIGRVTGVL 184

Query: 181 SSLLNIKVIMELTNHELVPIVKGRGLKTFSKWLDNFVESAQTRKIAEIGISYCGKADMAN 240
           SSLLN+KV+M L N EL  +VKGRG KTF+KWLD+++      R IAEI ISY G+A +A
Sbjct: 185 SSLLNVKVVMALKNDELKTLVKGRGNKTFTKWLDSYLAKNSHRPIAEIAISYAGEASLAL 244

Query: 241 NFREKLAV-LGAPISVLETGSIIQTHTGEDAFAVMVRYE 278
             +E++A      ISVLETGSIIQTHTGE AFAVMVRYE
Sbjct: 245 TLKERIAAYYNHSISVLETGSIIQTHTGEGAFAVMVRYE 283
```

SEQ ID 300 (GBS113) was expressed in *E.coli* as a His-fusion product. Purified protein is shown in Figure 201, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 90

A DNA sequence (GBSx0092) was identified in *S.agalactiae* <SEQ ID 307> which encodes the amino acid sequence <SEQ ID 308>. Analysis of this protein sequence reveals the following:

```
Possible site: 28

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA72097 GB:Y11213 hypothetical protein [Streptococcus thermophilus]
 Identities = 75/185 (40%), Positives = 116/185 (62%), Gaps = 3/185 (1%)

Query: 13  WKWAFLLLLAINLSFTAVIASRLIQVREPNTGKISTGVQDKVKVGTFTTNKSQLNKTIAL 72
```

```
            WKW FL LLA+NL+  +V+  R++   E +   +  G     K+G ++ +K +L++++
Sbjct: 5    WKWLFLGLLALNLALISVVTVRIMTPVETSPVSLPKGA---TKIGKYSMSKEELDESLRG 61

Query: 73   YLKQYQTKKMNYKIYAASSSILFEGSYQLLGYEVPLYIYFEPYRLTNGAVQLKVTSFSVG 132
            + + Y T KM +K+   +S I+FE SY++LG+ VPLY+YF P    +GAV L+ +  S G
Sbjct: 62   FAQDYSTDKMRFKVKVTNSKIVFESSYKVLGHAVPLYVYFTPLVSESGAVVLQESELSAG 121

Query: 133  TLPLPEKDVLQYIKSSYKLPNFVDIKPKKSVININLQDLKNKEGIYLKATAIDLVNDNFS 192
            TL LP  D L  IK S KLP+++ I  KK  + +N+Q +KN +GI  +A + DLVND
Sbjct: 122  TLKLPILDALNMIKRSTKLPDYIVIDSKKGKVILNIQSMKNDKGITARAQSFDLVNDRSE 181

Query: 193  FDIFK 197
            FDI+K
Sbjct: 182  FDIYK 186
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 309> which encodes the amino acid sequence <SEQ ID 310>. Analysis of this protein sequence reveals the following:

```
     Possible site: 29

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA72097 GB:Y11213 hypothetical protein [Streptococcus thermophilus]
 Identities = 73/185 (39%), Positives = 112/185 (60%), Gaps = 3/185 (1%)

Query: 10   WKWSFLCLLAFNTAFLMVIASRLIQVREPESELIAKKPVKNIKIGTFVTTREQLNETVAS 69
            WKW FL LLA N A + V+  R++   E     + K   K   IG +  ++E+L+E++
Sbjct: 5    WKWLFLGLLALNLALISVVTVRIMTPVETSPVSLPKGATK---IGKYSMSKEELDESLRG 61

Query: 70   YLKDYQTEKMSYKFYATSSSILFEGTYQLLGYEVPLYIYFQPHRLENGAVQLQVISFSVG 129
            + +DY T+KM +K   T+S I+FE +Y++LG+ VPLY+YF P   E+GAV LQ    S G
Sbjct: 62   FAQDYSTDKMRFKVKVTNSKIVFESSYKVLGHAVPLYVYFTPLVSESGAVVLQESELSAG 121

Query: 130  TLPLPEKDVLQYLKSSYKLPSFVKVMPNQSAIVVNLQDIQNDAKVYLKAKKIDLFNDEIS 189
            TL LP  D L  +K S KLP ++ +   +  +++N+Q ++ND  +  +A+  DL ND
Sbjct: 122  TLKLPILDALNMIKRSTKLPDYIVIDSKKGKVILNIQSMKNDKGITARAQSFDLVNDRSE 181

Query: 190  FNIYK 194
            F+IYK
Sbjct: 182  FDIYK 186
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 129/194 (66%), Positives = 155/194 (79%)

Query: 5    KTGRNLNFWKWAFLLLLAINLSFTAVIASRLIQVREPNTGKISTGVQDKVKVGTFTTNKS 64
            K   NLN+WKW+FL LLA N +F  VIASRLIQVREP +  I+      +K+GTF T +
Sbjct: 2    KKKSNLNWWKWSFLCLLAFNTAFLMVIASRLIQVREPESELIAKKPVKNIKIGTFVTTRE 61

Query: 65   QLNKTIALYLKQYQTKKMNYKIYAASSSILFEGSYQLLGYEVPLYIYFEPYRLTNGAVQL 124
            QLN+T+A YLK YQT+KM+YK YA SSSILFEG+YQLLGYEVPLYIYF+P+RL NGAVQL
Sbjct: 62   QLNETVASYLKDYQTEKMSYKFYATSSSILFEGTYQLLGYEVPLYIYFQPHRLENGAVQL 121

Query: 125  KVTSFSVGTLPLPEKDVLQYIKSSYKLPNFVDIKPKKSVININLQDLKNKEGIYLKATAI 184
            +V SFSVGTLPLPEKDVLQY+KSSYKLP+FV + P +S I +NLQD++N   +YLKA  I
Sbjct: 122  QVISFSVGTLPLPEKDVLQYLKSSYKLPSFVKVMPNQSAIVVNLQDIQNDAKVYLKAKKI 181

Query: 185  DLVNDNFSFDIFKK 198
            DL ND  SF+I+KK
Sbjct: 182  DLFNDEISFNIYKK 195
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8487> and protein <SEQ ID 8488> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 7
McG: Discrim Score:      7.47
GvH: Signal Score (-7.5): 2.42
     Possible site: 28
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:   5.89 threshold:  0.0
   PERIPHERAL  Likelihood =  5.89     120
 modified ALOM score:  -1.68

*** Reasoning Step: 3

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

SEQ ID 308 (GBS20) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 4 (lane 5; MW 25kDa) and in Figure 167 (lane 12-14; MW 37kDa – thioredoxin fusion). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 9 (lane 7; MW 47.6kDa). Purified Thio-GBS20-His is shown in Figure 244, lane 12.

Example 91

A DNA sequence (GBSx0093) was identified in *S.agalactiae* <SEQ ID 311> which encodes the amino acid sequence <SEQ ID 312>. This protein is predicted to be histone-like DNA-binding protein. Analysis of this protein sequence reveals the following:

```
Possible site: 40

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2768(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9313> which encodes amino acid sequence <SEQ ID 9314> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD40810 GB:L40355 histone-like DNA-binding protein [Streptococcus mutans]
 Identities = 43/47 (91%), Positives = 46/47 (97%)

Query: 1  MANKQDLIAKVAEATELTKKDSAAAVDAVFAAVADYLAEGEKVQLIG 47
          MANKQDLIAKVAEATELTKKDSAAAVDAVF+AV+ YLA+GEKVQLIG
Sbjct: 1  MANKQDLIAKVAEATELTKKDSAAAVDAVFSAVSSYLAKGEKVQLIG 47
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 313> which encodes the amino acid sequence <SEQ ID 314>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2834(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 41/47 (87%), Positives = 44/47 (93%)

Query: 1  MANKQDLIAKVAEATELTKKDSAAAVDAVFAAVADYLAEGEKVQLIG 47
          MANKQDLIAKVAEATELTKKDSAAAVDAVF+ +  +LAEGEKVQLIG
Sbjct: 1  MANKQDLIAKVAEATELTKKDSAAAVDAVFSTIEAFLAEGEKVQLIG 47
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 92

A DNA sequence (GBSx0094) was identified in *S.agalactiae* <SEQ ID 315> which encodes the amino acid sequence <SEQ ID 316>. Analysis of this protein sequence reveals the following:

```
Possible site: 54

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2722(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9293> which encodes amino acid sequence <SEQ ID 9294> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10793> which encodes amino acid sequence <SEQ ID 10794> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD17886 GB:AF100456 hyaluronate-associated protein precursor
            [Streptococcus equi]
 Identities = 303/435 (69%), Positives = 360/435 (82%), Gaps = 1/435 (0%)

Query: 1   MATKVDVSKDGLTYTATLRKGLKWSDGSKLTAKDFVYSWQRLVDPKTASQYAYLAVEGHV 60
           +A KVDVS+DGLTYTATLR GLKWSDGS LTA+DFVYSWQR+VDPKTAS+YAYLA E H+
Sbjct: 87  LAEKVDVSEDGLTYTATLRDGLKWSDGSDLTAEDFVYSWQRMVDPKTASEYAYLATESHL 146

Query: 61  LNADKINEGQEKDLNKLGVKAEGDDKVVITLSSPSPQFIYYLAFTNFMPQKQEVVEKYGK 120
            NA+ IN G+  DL+ LGVKA+G+ KV+ TL+ P+PQF   L+F+NF+PQK+  V+  GK
Sbjct: 147 KNAEDINSGKNPDLDSLGVKADGN-KVIFTLTEPAPQFKSLLSFSNFVPQKESFVKDAGK 205

Query: 121 DYATTSKNTVYSGPYTVEGWNGSNGTFTLKKNKNYWDAKNVKTKEVRIQTVKKPDTAVQM 180
           DY TTS+  +YSGPY V+ WNG++GTF L KNKNYWDAKNVKT+ V +QTVKKPDTAVQM
Sbjct: 206 DYGTTSEKQIYSGPYIVKDWNGTSGTFKLVKNKNYWDAKNVKTETVNVQTVKKPDTAVQM 265

Query: 181 YKRGELDAANISNTSAIYQANKNNKDVTDVLEATTAYMEYNTTGSVKGLDNVKIRRALNL 240
           YK+G+LD ANIS TSAIY ANK +KDV  VLEATTAY+ YN TG+++GL+++KIR+ALNL
Sbjct: 266 YKQGKLDFANISGTSAIYNANKKHKDVVPVLEATTAYIVYNQTGAIEGLNSLKIRQALNL 325

Query: 241 ATNRKGVVQAAVDTGSKPAIAFAPTGLAKTPDGTDLAKYVAPGYEYNKTEAAKLFKEGLA 300
           AT+RKG+V AAVDTGSKPA A  PTGLAK  DGTDL ++VAPGY+Y+  EAAKLFKEGLA
Sbjct: 326 ATDRKGIVSAAVDTGSKPATALVPTGLAKLSDGTDLTEHVAPGYKYDDKEAAKLFKEGLA 385

Query: 301 ESGLTKLKLTITADADAPAAKNSVDYIKSTWEAALPGLTVEEKFVTFKQRLEDSRKQNFD 360
           E G   L +TITADADAPAAK++VDYIK TWE ALPGLTVEEKFV FKQRLED++ QNF+
```

```
Sbjct: 386 ELGKDALTITITADADAPAAKSAVDYIKETWETALPGLTVEEKFVPFKQRLEDTKNQNFE 445

Query: 361 IVVSLWGGDYPEGSTFYGLFKSDSQNNDGKFANKDYDAAYNKAISEDAMKPAESAKDYKE 420
           + V LWGGDYP+GSTFYGLFKS S  N GKF N DYDAAYNKA++ DA+    +A DYK
Sbjct: 446 VAVVLWGGDYPKGSTFYGLFKSGSAYNYGKFTNADYDAAYNKALTTDALNTDAAADDYKA 505

Query: 421 AEKILFEQGAYNPLY 435
           AEK L++   YNPLY
Sbjct: 506 AEKALYDNALYNPLY 520
```

A related GBS gene <SEQ ID 8489> and protein <SEQ ID 8490> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: 21   Crend: 4
       Sequence Pattern: CGSK
SRCFLG: 0
McG: Length  of  UR:   19
     Peak Value of UR:   2.34
     Net Charge of CR: 3
McG: Discrim Score:      5.94
GvH: Signal Score (-7.5): 0.6
     Possible site: 20
>>> May be a lipoprotein
Amino Acid Composition: calculated from 22
ALOM program   count: 0 value:   5.14 threshold:  0.0
   PERIPHERAL  Likelihood =  5.14     166
 modified ALOM score:  -1.53

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP|4336671|gb|AAD17886.1||AF100456 hyaluronate-associated protein
           precursor {Streptococcus equi}

 Score =  721 bits (1840), Expect = 0.0
 Identities = 354/515 (68%), Positives = 417/515 (80%), Gaps = 2/515 (0%)

Query: 1   KNWRRVGVGVLTLASVATLAACGSK-SASQDSNGAINWAIPTEINTLDLSKVTDTYSNLA 59
           K  +R+G+  +TLASVA L ACG+K SAS D    INW  PTEI TLD+SK TDTYS LA
Sbjct: 7   KACKRLGLAAVTLASVAALMACGNKQSASTDKKSEINWYTPTEIITLDISKNTDTYSALA 66

Query: 60  IGNSSSNFLRLDKDGKTRPDLATKVDVSKDGLTYTATLRKGLKWSDGSKLTAKDFVYSWQ 119
           IGNS SN LR D  GK +PDLA KVDVS+DGLTYTATLR GLKWSDGS LTA+DFVYSWQ
Sbjct: 67  IGNSGSNLLRADAKGKLQPDLAEKVDVSEDGLTYTATLRDGLKWSDGSDLTAEDFVYSWQ 126

Query: 120 RLVDPKTASQYAYLAVEGHVLNADKINEGQEKDLNKLGVKAEGDDKVVITLSSPSPQFIY 179
           R+VDPKTAS+YAYLA E H+ NA+ IN G+  DL+ LGVKA+G+ KV+ TL+ P+PQF
Sbjct: 127 RMVDPKTASEYAYLATESHLKNAEDINSGKNPDLDSLGVKADGN-KVIFTLTEPAPQFKS 185

Query: 180 YLAFTNFMPQKQEVVEKYGKDYATTSKNTVYSGPYTVEGWNGSNGTFTLKKNKNYWDAKN 239
            L+F+NF+PQK+  V+  GKDY TTS+  +YSGPY V+ WNG++GTF L KNKNYWDAKN
Sbjct: 186 LLSFSNFVPQKESFVKDAGKDYGTTSEKQIYSGPYIVKDWNGTSGTFKLVKNKNYWDAKN 245

Query: 240 VKTKEVRIQTVKKPDTAVQMYKRGELDAANISNTSAIYQANKNNKDVTDVLEATTAYMEY 299
           VKT+ V +QTVKKPDTAVQMYK+G+LD ANIS TSAIY ANK +KDV  VLEATTAY+ Y
Sbjct: 246 VKTETVNVQTVKKPDTAVQMYKQGKLDFANISGTSAIYNANKKHKDVVPVLEATTAYIVY 305

Query: 300 NTTGSVKGLDNVKIRRALNLATNRKGVVQAAVDTGSKPAIAFAPTGLAKTPDGTDLAKYV 359
           N TG+++GL+++KIR+ALNLAT+RKG+V AAVDTGSKPA A  PTGLAK  DGTDL ++V
Sbjct: 306 NQTGAIEGLNSLKIRQALNLATDRKGIVSAAVDTGSKPATALVPTGLAKLSDGTDLTEHV 365
```

```
Query: 360 APGYEYNKTEAAKLFKEGLAESGLTKLKLTITADADAPAAKNSVDYIKSTWEAALPGLTV 419
           APGY+Y+  EAAKLFKEGLAE G   L +TITADADAPAAK++VDYIK TWE ALPGLTV
Sbjct: 366 APGYKYDDKEAAKLFKEGLAELGKDALTITITADADAPAAKSAVDYIKETWETALPGLTV 425

Query: 420 EEKFVTFKQRLEDSRKQNFDIVVSLWGGDYPEGSTFYGLFKSDSQNNDGKFANKDYDAAY 479
           EEKFV FKQRLED++ QNF++ V LWGGDYP+GSTFYGLFKS S  N GKF N DYDAAY
Sbjct: 426 EEKFVPFKQRLEDTKNQNFEVAVVLWGGDYPKGSTFYGLFKSGSAYNYGKFTNADYDAAY 485

Query: 480 NKAISEDAMKPAESAKDYKEAEKILFEQGAYNPLY 514
           NKA++ DA+    +A DYK AEK L++   YNPLY
Sbjct: 486 NKALTTDALNTDAAADDYKAAEKALYDNALYNPLY 520
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 317> which encodes the amino acid sequence <SEQ ID 318>. Analysis of this protein sequence reveals the following:

```
      Possible site: 24

>>> May be a lipoprotein

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 114/428 (26%), Positives = 185/428 (42%), Gaps = 63/428 (14%)

Query: 7   VSKDGLTYTATLRKGLKW--SDGSK---LTAKDFVYSWQRLVDPKTASQYAYLAVEGHVL 61
           VSKDGLTYT TLR G+ W  +DG +    +TA+DFV   +  VD K+ + Y    VE  +
Sbjct: 92  VSKDGLTYTYTLRDGVSWYTADGEEYAPVTAEDFVTGLKHAVDDKSDALY---VVEDSIK 148

Query: 62  NADKINEGQEKDLNKLGVKAEGDDKVVITLSSPSPQFIYYLAFTNFMPQKQEVVEKYGKD 121
           N      G E D  ++GVKA  D  V  TL+ P   +     ++   P   + ++  GKD
Sbjct: 149 NLKAYQNG-EVDFKEVGVKALDDKTVQYTLNKPESYWNSKTTYSVLFPVNAKFLKSKGKD 207

Query: 122 YATTSKNTV-YSGPYTVEGWNGSNGTFTLKKNKNYWDAKNVKTKEVRI--QTVKKPDTAV 178
           + TT  +++  +G Y +  +  S  +    KN+NYWDAKNV  + V++       P +
Sbjct: 208 FGTTDPSSILVNGAYFLSAFT-SKSSMEFHKNENYWDAKNVGIESVKLTYSDGSDPGSFY 266

Query: 179 QMYKRGELDAANISNTSAIYQANKNN--KDVT-DVLEATTAYMEYNTT------------ 223
           + + +GE   A +      Y++ K N   ++T  +L     ++ +N
Sbjct: 267 KNFDKGEFSVARLYPNDPTYKSAKKNYADNITYGMLTGDIRHLTWNLNRTSFKNTKKDPA 326

Query: 224 ---GSVKGLDNVKIRRALNLATNRKGVVQAAVDTGSKPA----IAFAPT--GLAKTPDGT 274
                 K L+N   R+A+  A +R            +K      +   PT   + ++  G+
Sbjct: 327 QQDAGKKALNNKDFRQAIQFAFDRASFQAQTAGQDAKTKALRNMLVPPTFVTIGESDFGS 386

Query: 275 DLAKYVAP-GYE-------------YNKTEAAKLF---KEGLAESGLT-KLKLTITADAD 316
           ++ K +A  G E              YN  +A   F   KE L   G+T  ++L    D
Sbjct: 387 EVEKEMAKLGDEWKDVNLADAQDGFYNPEKAKAEFAKAKEALTAEGVTFPVQLDYPVDQA 446

Query: 317 APAAKNSVDYIKSTWEAALPGLTV-----EEKFVTFKQR---LEDSRKQNFDIVVSLWGG 368
             A         K + EA+L    V     E +  T + +    E   +Q++DI+ S WG
Sbjct: 447 NAATVQEAQSFKQSVEASLGKENVIVNVLETETSTHEAQGFYAETPEQQDYDIISSWWGP 506

Query: 369 DYPEGSTF 376
           DY +  T+
Sbjct: 507 DYQDPRTY 514
```

SEQ ID 9294 (GBS663) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 137 (lane 3; MW 89.5kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 137 (lane 5-7; MW 64.5kDa), in Figure

179 (lane 11; MW 65kDa) and in Figure 65 (lane 2; MW 61kDa). Purified GBS663-His is shown in Figure 231, lane 3-4. Purified GBS324-His is shown in lane 6 of Figure 210.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 93

A DNA sequence (GBSx0095) was identified in *S.agalactiae* <SEQ ID 319> which encodes the amino acid sequence <SEQ ID 320>. This protein is predicted to be transmembrane protein OppB (oppB). Analysis of this protein sequence reveals the following:

```
Possible site: 37

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.77   Transmembrane  293 - 309 ( 281 - 313)
   INTEGRAL    Likelihood = -9.77   Transmembrane   21 -  37 (  14 -  46)
   INTEGRAL    Likelihood = -6.32   Transmembrane  115 - 131 ( 105 - 132)
   INTEGRAL    Likelihood = -4.88   Transmembrane  144 - 160 ( 140 - 166)
   INTEGRAL    Likelihood = -3.03   Transmembrane  238 - 254 ( 237 - 255)

----- Final Results -----
               bacterial membrane --- Certainty=0.5310(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8491> which encodes amino acid sequence <SEQ ID 8492> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF73091 GB:AF103793 transmembrane protein OppB [Listeria monocytogenes]
 Identities = 147/304 (48%), Positives = 221/304 (72%), Gaps = 1/304 (0%)

Query: 13  MIKYILKRVAILLVTLWVVITLSFFLMQILPGTPYNNP-KLTEEMIALLNKQYGLDKPVW 71
           M+KY LKRV  +L+TL+++ +++F LM+ LPGTPY N  KL++E I + N++YGL+  +
Sbjct: 1   MVKYTLKRVLYMLITLFIIASVTFVLMKFLPGTPYRNQEKLSDEQIHMTNEKYGLNDSIP 60

Query: 72  QQYLTYLWNVLHGDFGTSYQSVNQPVSRMISLRLGVSVHLGVQALVFGVLGGILVGAISA 131
            QY  Y+  ++ GD G S+Q  N+PVS ++S  +G SV L ++A+ FGV+ GIL+G I+A
Sbjct: 61  VQYFNYMTGLVKGDLGVSFQLDNRPVSEILSALIGPSVQLALEAMAFGVIFGILLGVIAA 120

Query: 132 RHKNDKVDGILSVIATLGISMPSFIIGILLLDYFGFKWNLLPLSGWGTFSQTILPSLALG 191
            ++N   D   + IA LG S+PSF+   +L  + G K  + P++GWGTF+ TILP+ AL
Sbjct: 121 MYQNRWPDYTSTFIAILGKSVPSFVFATVLQYWLGAKLQIFPVAGWGTFADTILPAFALA 180

Query: 192 LPTLASVSRFFRSEMIETLNSDYVQLARSKGMTIRQVTRKHAYRNSMIPILTLIGPLAAG 251
           +  LA+ +RF R+E+I+   SDYV LA++KG +  +V  KHA RN++IP++T++GPL+
Sbjct: 181 MFPLATAARFMRTELIDVFASDYVLLAKAKGNSRTEVAVKHAIRNALIPLITVLGPLSVA 240

Query: 252 LLTGSALIEQIFSIPGIGQQFVTSIPTKDYPVIMGTTIVYAVMLMVAILITDVVISIVDP 311
           L+TGS +IE I+SIPGIG QFV+SI T DYPVIMGTTI++AVML+  IL+ D++  ++DP
Sbjct: 241 LMTGSLVIENIYSIPGIGSQFVSSIQTNDYPVIMGTTILFAVMLVFVILVVDILYGLIDP 300

Query: 312 RVRL 315
           R+R+
Sbjct: 301 RIRV 304
```

There is also homology to SEQ ID 64.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 9069> which encodes amino acid sequence <SEQ ID 9070>. Analysis of this protein sequence reveals the following:

```
    Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.81   Transmembrane  466 - 482 ( 463 - 493)
   INTEGRAL    Likelihood = -5.10   Transmembrane  419 - 435 ( 418 - 440)
   INTEGRAL    Likelihood = -4.78   Transmembrane  328 - 344 ( 322 - 348)
   INTEGRAL    Likelihood = -4.41   Transmembrane  366 - 382 ( 365 - 384)
   INTEGRAL    Likelihood = -4.09   Transmembrane  290 - 306 ( 287 - 311)
   INTEGRAL    Likelihood = -2.97   Transmembrane   17 -  33 (  13 -  36)

----- Final Results -----
               bacterial membrane --- Certainty=0.4524(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS sequences follows:

```
 Score =  117 bits (291), Expect = 3e-28
 Identities = 61/208 (29%), Positives = 121/208 (57%), Gaps = 4/208 (1%)

Query: 291 IGFFGVMFSYIVGLPLGLFMARFKNTYFDSFSTATMTFMLALPSIAV-IYVVRFLGGMVG 349
           +G   ++F  + G+ +G   AR KN   D   +   T  +++PS  + I ++ + G
Sbjct: 99  LGVQALVFGVLGGILVGAISARHKNDKVDGILSVIATLGISMPSFIIGILLLDYFGFKWN 158

Query: 350 LPDSFPMLGASDPKSYILPALILGILNIPTTVIWFRRYLVDLQASDWVRFARSKGLSESE 409
           L    P+ G       ILP+L LG+  + +   +FR  +++   SD+V+ ARSKG++  +
Sbjct: 159 L---LPLSGWGTFSQTILPSLALGLPTLASVSRFFRSEMIETLNSDYVQLARSKGMTIRQ 215

Query: 410 IYRGHLFKNAMVPIVSGVPASIILAIGGATLTETVFAFPGMGKMLIDSIKSANNSMIVGL 469
           + R H ++N+M+PI++ +       + G+ L E +F+ PG+G+  + SI + +  +I+G
Sbjct: 216 VTRKHAYRNSMIPILTLIGPLAAGLLTGSALIEQIFSIPGIGQQFVTSIPTKDYPVIMGT 275

Query: 470 TFIFTVLSIVSLLLGDIVMTLVDPRIKL 497
           T ++ V+ +V++L+ D+V+++VDPR++L
Sbjct: 276 TIVYAVMLMVAILITDVVISIVDPRVRL 303
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 94

A DNA sequence (GBSx0096) was identified in *S.agalactiae* <SEQ ID 321> which encodes the amino acid sequence <SEQ ID 322>. This protein is predicted to be transmembrane protein OppC (oppC). Analysis of this protein sequence reveals the following:

```
Possible site: 59

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.52   Transmembrane  311 - 327 ( 307 - 333)
   INTEGRAL    Likelihood = -7.80   Transmembrane   42 -  58 (  40 -  65)
   INTEGRAL    Likelihood = -7.43   Transmembrane  142 - 158 ( 131 - 165)
   INTEGRAL    Likelihood = -4.73   Transmembrane  182 - 198 ( 179 - 214)
   INTEGRAL    Likelihood = -3.50   Transmembrane  257 - 273 ( 257 - 276)

----- Final Results -----
               bacterial membrane --- Certainty=0.5607(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF73092 GB:AF103793 transmembrane protein OppC [Listeria
           monocytogenes]
 Identities = 157/325 (48%), Positives = 219/325 (67%), Gaps = 4/325 (1%)

Query: 20  EKIEKPALSFMQDAWRRLKKNKLAVVSLYLLALLLTFSLASNLFVTQKDANGFDSKKVTT 79
```

```
           EKI +P+L+F+QD+W R++KNK A+VSL +LAL++  ++                ++++T
Sbjct: 22  EKINRPSLTFLQDSWLRIRKNKAALVSLIVLALVIIMAIVGPYLSQNLGPEHNINRQITE 81

Query: 80  YRNLPPKLSS--NLPFWNGSIKYAGNTESTDAYKSQNVPEKVKYALGTDSLGRSVAKRII 137
             +LPPK+    N+PFWNG     G  E  D YK  N+ E   Y LG+D+LGR    RI
Sbjct: 82  NASLPPKVQGFENMPFWNGHQSIGG--EDVDIYKQNNIKEGTYYWLGSDTLGRDQFARIW 139

Query: 138 VGIRISLLVAIAATFIDLIIGVTYGLVSGFAGGRLDTLMQRIVEVISSIPNLVIVTMLGL 197
            G R+SL++A+ A   DL+IGV YGL+SG+ GGR+D  MQR++EVI +IPNLV+V ++ L
Sbjct: 140 AGTRVSLIIAVVAALCDLVIGVAYGLISGYVGGRVDNFMQRVLEVIGAIPNLVVVILMML 199

Query: 198 VLGNGITAIIISIAFTGWTSMSRQVRNLTLSYREREFVLAARSLGESPIKIAFKHILPNI 257
           +L  GI +III+IA T W +M+R VR   L  + +EFV+A+ +LGES  KI  KH++PNI
Sbjct: 200 ILEPGIVSIIIAIAMTSWITMARVVRGQVLKRKNQEFVMASMTLGESTPKILIKHLIPNI 259

Query: 258 SGIIIVQIMMTIPSAIMYEAVLSAINLGVKPPTASLGSLISDAQENLQYYPYQVILPALA 317
           SGIII+ IM +IPSAI +EA LS I LG+  P ASLG L++D  + LQ  PY ++ P +
Sbjct: 260 SGIIIINIMFSIPSAIFFEAFLSFIGLGLPAPAASLGVLVNDGYKTLQVLPYMILYPCIV 319

Query: 318 LVMISLAFILLGDGLRDAFDPKSSD 342
           L +I +AF L+ DGLRDAFDPK  D
Sbjct: 320 LCIIMIAFNLIADGLRDAFDPKMRD 344
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 323> which encodes the amino acid sequence <SEQ ID 324>. Analysis of this protein sequence reveals the following:

```
Possible site: 59

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.30   Transmembrane   43 -  59 (  37 -  65)
   INTEGRAL    Likelihood = -8.49   Transmembrane  111 - 127 ( 109 - 135)
   INTEGRAL    Likelihood = -6.26   Transmembrane  279 - 295 ( 270 - 298)
   INTEGRAL    Likelihood = -3.88   Transmembrane  172 - 188 ( 172 - 188)
   INTEGRAL    Likelihood = -3.61   Transmembrane  145 - 161 ( 145 - 165)
   INTEGRAL    Likelihood = -1.49   Transmembrane  223 - 239 ( 223 - 239)

----- Final Results -----
              bacterial membrane --- Certainty=0.5118(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 91/325 (28%), Positives = 156/325 (48%), Gaps = 34/325 (10%)

Query: 16  SSTQEKIEKPALSFMQDAWRRLKKNKLAVVSLYLLALLLTFSLASNLFVTQKDANGFDSK 75
           S   E I+ PA S+ +  +R+    K  V  L +L  +L  S    +F         +D
Sbjct: 16  SEASEVIDTPAYSYWKSVFRQFFSKKSTVFMLVILVTVLMMSFIYPMFAN------YDFN 69

Query: 76  KVTTYRNLPPKLSSNLPFWNGSIKYAGNTESTDAYKSQNVPEKVKYALGTDSLGRSVAKR 135
            V+   +                          +  + +    +Y  GTD  G+S+
Sbjct: 70  DVSNIND--------------------------FSKRYIWPNAEYWFGTDKNGQSLFDG 102

Query: 136 IIVGIRISLLVAIAATFIDLIIGVTYGLVSGFAGGRLDTLMQRIVEVISSIPNLVIVTML 195
           +  G R S+L+++ AT I++ IGV  G + G +    D +M  I  +IS+IP+++I+ +L
Sbjct: 103 VWYGARNSILISVIATLINITIGVVLGAIWGVSKA-FDKVMIEIYNIISNIPSMLIIIVL 161

Query: 196 GLVLGNGITAIIISIAFTGWTSMSRQVRNLTLSYREREFVLAARSLGESPIKIAFKHILP 255
              LG G   +I++   TGW  ++  +R   L YR+ E+ LA+++LG    KIA K++LP
Sbjct: 162 TYSLGAGFWNLILAFCITGWIGVAYSIRVQILRYRDLEYNLASQTLGTPMYKIAVKNLLP 221

Query: 256 NISGIIIVQIMMTIPSAIMYEAVLSAINLGVKPPTASLGSLISDAQENLQYYPYQVILPA 315
            +  +I+  +   +P  +  EA LS   +G+   T SLG  I++   NL    Y   +P
Sbjct: 222 QLVSVIMTMLSQMLPVYVSSEAFLSFFGIGLPTTTPSLGRFIANYSSNLTTNAYLFWIPL 281

Query: 316 LALVMISLAFILLGDGLRDAFDPKS 340
           + L+++SL   ++G  L DA DP+S
```

```
Sbjct: 282 VTLILVSLPLYIVGQNLADASDPRS 306
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 95

A DNA sequence (GBSx0097) was identified in *S.agalactiae* <SEQ ID 325> which encodes the amino acid sequence <SEQ ID 326>. This protein is predicted to be ATPase OppD (oppD). Analysis of this protein sequence reveals the following:

```
Possible site: 20

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.85   Transmembrane  164 - 180 ( 163 - 180)

----- Final Results -----
               bacterial membrane --- Certainty=0.1341(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF73093 GB:AF103793 ATPase OppD [Listeria monocytogenes]

Identities = 230/342 (67%), Positives = 283/342 (82%), Gaps = 2/342 (0%)

Query: 4   ETILSVNNLHVDFHTYAGEVKAIRDVNFELKKGETLAIVGESGSGKSVTTRTLIGLNAK- 62
           E +L V +L++ FHTYAGEVKAIR VNF+L KGETLAIVGESGSGKSVTT++++ L  +
Sbjct: 2   EKLLEVKDLNISFHTYAGEVKAIRGVNFDLYKGETLAIVGESGSGKSVTTKSIMRLLPEG 61

Query: 63  NSEI-SGNVQFKGRNLVELSEEEWTKVRGNEISMIFQDPMTSLDPTMKIGMQIAEPMMIH 121
           NSEI SG + F G ++ +  E++  K+RG +I+MIFQDPMTSL+PTM IG QI+EP++ H
Sbjct: 62  NSEIKSGQILFNGMDIAKAHEKQMQKIRGKDIAMIFQDPMTSLNPTMTIGKQISEPLIKH 121

Query: 122 QKISKKDALKLALELMKDVGIPNAEEHINDYPHQWSGGMRQRAVIAIALAADPEILIADE 181
           QKISK +A K AL L++ VGI NAEE I  YPHQ+SGGMRQR VIAI+LA +P+ILIADE
Sbjct: 122 QKISKHEAHKTALRLLQLVGIANAEERIKQYPHQFSGGMRQRVVIAISLACNPQILIADE 181

Query: 182 PTTALDVTIQAQILNLMKKIQAERDSSIVFITHDLGVVAGMADRVAVMYAGKIVEFGTVD 241
           PTTALDVTIQAQIL+LMK +Q + D+SI+FITHDLGVVA +ADRVAVMY GKIVE GTVD
Sbjct: 182 PTTALDVTIQAQILDLMKDLQKKIDTSIIFITHDLGVVANVADRVAVMYGGKIVEIGTVD 241

Query: 242 EVFYNPQHPYTWGLLNSMPTTDTESGSLESIPGTPPDLLNPPKGDAFAARNEFALDIDHE 301
           E+FYNPQHPYTWGL++SMPT DT+   L  IPGTPPDLL+PPKGDAFAARN++A+ ID E
Sbjct: 242 EIFYNPQHPYTWGLISSMPTLDTDDEELFVIPGTPPDLLHPPKGDAFAARNKYAMQIDLE 301

Query: 302 EEPPYFKVSETHFAATWLLDERSPKVLPPLPIQKRWEKWNEI 343
           EEPP FKVS+TH+AATWLL   +P+V PP  + +R E++ E+
Sbjct: 302 EEPPLFKVSDTHYAATWLLHPDAPEVTPPDAVLRRQEQFAEL 343
```

There is also homology to SEQ ID 72.

SEQ ID 326 (GBS375) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 64 (lane 9; MW 42kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 71 (lane 3; MW 67kDa).

GBS375-GST was purified as shown in Figure 215, lane 10.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 96

A DNA sequence (GBSx0098) was identified in *S.agalactiae* <SEQ ID 327> which encodes the amino acid sequence <SEQ ID 328>. Analysis of this protein sequence reveals the following:

```
Possible site: 28

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3060(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA62692 GB:M57689 sporulation protein [Bacillus subtilis]
 Identities = 195/308 (63%), Positives = 245/308 (79%), Gaps = 4/308 (1%)

Query: 1   MTENRKKLVEVKNVSLTFNKGKANEVRAIDNVSFDIYEGEVFGLVGESGSGKTTVGRSIL 60
           M E  +KL+E+K++   F   +   V+A+D++SFDIY+GE  GLVGESG GK+T GRSI+
Sbjct: 1   MNELTEKLLEIKHLKQHFVTPRGT-VKAVDDLSFDIYKGETLGLVGESGCGKSTTGRSII 59

Query: 61  KLYDISDGEITFNGEVISHLKG-KALHSFRKDAQMIFQDPQASLNGRMKIRDIVAEGLDI 119
           +LY+ +DGE+ FNGE +   K  K L  F +  QMIFQDP ASLN RM + DI+AEGLDI
Sbjct: 60  RLYEATDGEVLFNGENVHGRKSRKKLLEFNRKMQMIFQDPYASLNPRMTVADIIAEGLDI 119

Query: 120 HKLAKSKSDRDSKVQALLDLVGLNKDHLTRYPHEFSGGQRQRIGIARALAVEPKFIIADE 179
           HKLAK+K +R  +V  LL+ VGLNK+H  RYPHEFSGGQRQRIGIARALAV+P+FIIADE
Sbjct: 120 HKLAKTKKERMQRVHELLETVGLNKEHANRYPHEFSGGQRQRIGIARALAVDPEFIIADE 179

Query: 180 PISALDVSIQAQVVNLMQKLQREQGLTYLFIAHDLSMVKYISDRIGVMHWGKLLEVGTSD 239
           PISALDVSIQAQVVNLM++LQ+E+GLTYLFIAHDLSMVKYISDRIGVM++GKL+E+  +D
Sbjct: 180 PISALDVSIQAQVVNLMKELQKEKGLTYLFIAHDLSMVKYISDRIGVMYFGKLVELAPAD 239

Query: 240 DVYNNPIHPYTKSLLSAIPEPDPESERQRVHQPYNPAIEQ--DGQERQMHEITPGHFVLS 297
           ++Y NP+HPYTKSLLSAIP PDP+ ER RV Q Y+P++ Q  DG+  +  E+ PGHFV+
Sbjct: 240 ELYENPLHPYTKSLLSAIPLPDPDYERNRVRQKYDPSVHQLKDGETMEFREVKPGHFVMC 299

Query: 298 TPQEAEEY 305
           T  E + +
Sbjct: 300 TEAEFKAF 307
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 329> which encodes the amino acid sequence <SEQ ID 330>. Analysis of this protein sequence reveals the following:

```
Possible site: 47

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3900(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 164/306 (53%), Positives = 228/306 (73%), Gaps = 3/306 (0%)

Query: 6   KKLVEVKNVSLTFNKGKANEVRAIDNVSFDIYEGEVFGLVGESGSGKTTVGRSILKLYDI 65
           +KLVEVK++ ++F +GK   V A+ N +F I +GE F LVGESGSGKTT+GR+I+ L D
Sbjct: 3   EKLVEVKDLEISFGEGKKKFV-AVKNANFFIKKGETFSLVGESGSGKTTIGRAIIGLNDT 61

Query: 66  SDGEITFNGEVISHLKGKA-LHSFRKDAQMIFQDPQASLNGRMKIRDIVAEGLDIHKLAK 124
           S G+I ++G+VI+  K K+  +   +  QMIFQDP ASLN R  +  I++EGL    L K
Sbjct: 62  SSGQILYDGKVINGRKSKSEANELIRKIQMIFQDPAASLNERATVDYIISEGLYNFNLFK 121
```

```
Query: 125 SKSDRDSKVQALLDLVGLNKDHLTRYPHEFSGGQRQRIGIARALAVEPKFIIADEPISAL 184
           ++ +R  K++ ++  VGL  +HLTRYPHEFSGGQRQRIGIARAL + P+F+IADEPISAL
Sbjct: 122 TEEERKEKIKNMMAEVGLLSEHLTRYPHEFSGGQRQRIGIARALVMNPEFVIADEPISAL 181

Query: 185 DVSIQAQVVNLMQKLQREQGLTYLFIAHDLSMVKYISDRIGVMHWGKLLEVGTSDDVYNN 244
           DVS++AQV+NL++++Q E+GLTYLFIAHDLS+V++ISDRI V+H G ++EV  +++++NN
Sbjct: 182 DVSVRAQVLNLLKRMQAEKGLTYLFIAHDLSVVRFISDRIAVIHKGVIVEVAETEELFNN 241

Query: 245 PIHPYTKSLLSAIPEPDPESERQRVHQPYNPAIEQDGQER-QMHEITPGHFVLSTPQEAE 303
           PIHPYT+SLLSA+P PDP  ERQ+    Y+P       ++  M EI P HFV +   E E
Sbjct: 242 PIHPYTQSLLSAVPIPDPILERQKELVVYHPDQHDYTLDKPSMVEIKPNHFVWANQAEIE 301

Query: 304 EYKKQI 309
           +Y+K++
Sbjct: 302 KYQKEL 307
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 97

A repeated DNA sequence (GBSx0099) was identified in *S.agalactiae* <SEQ ID 331> which encodes the amino acid sequence <SEQ ID 332>. Analysis of this protein sequence reveals the following:

```
Possible site: 28

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3021(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 98

A repeated DNA sequence (GBSx0100) was identified in *S.agalactiae* <SEQ ID 333> which encodes the amino acid sequence <SEQ ID 334>. Analysis of this protein sequence reveals the following:

```
Possible site: 24

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0352(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 99

A repeated DNA sequence (GBSx0101) was identified in *S.agalactiae* <SEQ ID 335> which encodes the amino acid sequence <SEQ ID 336>. Analysis of this protein sequence reveals the following:

```
Possible site: 23

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5857(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 100

A repeated DNA sequence (GBSx0103) was identified in *S.agalactiae* <SEQ ID 337> which encodes the amino acid sequence <SEQ ID 338>. Analysis of this protein sequence reveals the following:

```
Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1472(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 101

A repeated DNA sequence (GBSx0104) was identified in *S.agalactiae* <SEQ ID 339> which encodes the amino acid sequence <SEQ ID 340>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0111(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 102

A repeated DNA sequence (GBSx0105) was identified in *S.agalactiae* <SEQ ID 341> which encodes the amino acid sequence <SEQ ID 342>. Analysis of this protein sequence reveals the following:

```
Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5628(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 103

A repeated DNA sequence (GBSx0106) was identified in *S.agalactiae* <SEQ ID 343> which encodes the amino acid sequence <SEQ ID 344>. Analysis of this protein sequence reveals the following:

```
Possible site: 39

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2059(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 104

A repeated DNA sequence (GBSx0107) was identified in *S.agalactiae* <SEQ ID 345> which encodes the amino acid sequence <SEQ ID 346>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2045(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 105

A DNA sequence (GBSx0108) was identified in *S.agalactiae* <SEQ ID 347> which encodes the amino acid sequence <SEQ ID 348>. Analysis of this protein sequence reveals the following:

```
Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3031(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB11822 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 125/282 (44%), Positives = 184/282 (64%)

Query: 1   MKIFEKAPAKLNLGLDIKGRCDDGYHELAMIMVSIDLNDYVTISELKEDCIVIDSDSSKM 60
           M+I EKAPAK+NL LD+  +  DGYHE+ MIM +IDL D + ++EL ED + + S +  +
Sbjct: 1   MRILEKAPAKINLSLDVTRKRPDGYHEVEMIMTTIDLADRIELTELAEDEVRVSSHNRFV 60

Query: 61  PLNNDNDVFKAADIIKNQYGINKGVHIRLEKSIPVCAGLGGGSTDAAATIRALNRLWNLQ 120
           P +  N  ++AA +IK++Y + KGV I + K IPV AGL GGS+DAAAT+R LNRLWNL
Sbjct: 61  PDDQRNLAYQAAKLIKDRYNVKKGVSIMITKVIPVAAGLAGGSSDAAATLRGLNRLWNLN 120

Query: 121 MDYDEMVAIGFKIGSDVPYCLGGGCSLVLGKGEIVKPLPTLRPCWIVLVKPDFGISTKSI 180
           +  + +  +G +IGSDV +C+ GG +L  G+GE +K + T   CW++L KP  G+ST  +
Sbjct: 121 LSAETLAELGAEIGSDVSFCVYGGTALATGRGEKIKHISTPPHCWVILAKPTIGVSTAEV 180

Query: 181 FRDIDCKSISRVDIDLLKSAILSSDYQLMVKSMGNSLEDITITKNPVISTIKERMLNSGA 240
           +R +    I   D+  +  AI    +Q M   +GN LE +T+  +P ++ IK +M   GA
Sbjct: 181 YRALKLDGIEHPDVQGMIEAIEEKSFQKMCSRLGNVLESVTLDMHPEVAMIKNQMKRFGA 240

Query: 241 DVALMTGSGPTVFSMCSTEKKADRVFNSMKGFCKEVYKVRLL 282
           D  LM+GSGPTVF +   E K  R++N ++GFC +VY VR++
Sbjct: 241 DAVLMSGSGPTVFGLVQYESKVQRIYNGLRGFCDQVYAVRMI 282
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 349> which encodes the amino acid sequence <SEQ ID 350>. Analysis of this protein sequence reveals the following:

```
Possible site: 44

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.87   Transmembrane   28 -  44 (  27 -  45)

----- Final Results -----
               bacterial membrane --- Certainty=0.2147(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 33/52 (63%), Positives = 38/52 (72%)

Query: 126 MVAIGFKIGSDVPYCLGGGCSLVLGKGEIVKPLPTLRPCWIVLVKPDFGIST 177
           M+ IG  IGSDVPYCL  GC+ V GKGE+V  +  L   W+VLVKPDFGIST
Sbjct: 1   MMDIGIPIGSDVPYCLLSGCAQVTGKGEVVCRILGLLSSWVVLVKPDFGIST 52
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 106

A DNA sequence (GBSx0109) was identified in *S.agalactiae* <SEQ ID 351> which encodes the amino acid sequence <SEQ ID 352>. This protein is predicted to be AdcR protein. Analysis of this protein sequence reveals the following:

```
Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1264(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA96184 GB:Z71552 AdcR protein [Streptococcus pneumoniae]
 Identities = 77/146 (52%), Positives = 117/146 (79%)

Query: 1   MTVLEQKLDHLVSQILLKAENQHELLFGTCQSDVKLTNTQEHILMLLSQEQLTNSDLAKK 60
           M  L + ++  +++++L+AENQHE+L G C S+V LTNTQEHILMLLS+E LTNS+LA++
Sbjct: 1   MRQLAKDINAFLNEVILQAENQHEILIGHCTSEVALTNTQEHILMLLSEESLTNSELARR 60

Query: 61  LNISQAAVTKAVKSLISQDMLKANKDSKDARITYFELSELAKPIADEHTHHHDNTLGVYG 120
           LN+SQAAVTKA+KSL+ + ML+ +KDSKDAR+ +++L++LA+PIA+EH HHH++TL  Y
Sbjct: 61  LNVSQAAVTKAIKSLVKEGMLETSKDSKDARVIFYQLTDLARPIAEEHHHHHHEHTLLTYE 120

Query: 121 RLVNHFSKDEKVVLERFLDLFSRELE 146
           ++   F+ +E+ V++RFL     E++
Sbjct: 121 QVATQFTPNEQKVIQRFLTALVGEIK 146
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 353> which encodes the amino acid sequence <SEQ ID 354>. Analysis of this protein sequence reveals the following:

```
Possible site: 28

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1536(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 106/147 (72%), Positives = 126/147 (85%)

Query: 1   MTVLEQKLDHLVSQILLKAENQHELLFGTCQSDVKLTNTQEHILMLLSQEQLTNSDLAKK 60
           M +LE+KLD+LV+ ILLKAENQHELLFG CQSDVKLTNTQEHILMLLSQ++LTN+DLAK
Sbjct: 1   MGILEKKLDNLVNTILLKAENQHELLFGACQSDVKLTNTQEHILMLLSQQRLTNTDLAKA 60

Query: 61  LNISQAAVTKAVKSLISQDMLKANKDSKDARITYFELSELAKPIADEHTHHHDNTLGVYG 120
           LNISQAAVTKA+KSL+ QDML   KD+ DAR+TYFEL+ELAKPIA EHTHHHD TL VY
Sbjct: 61  LNISQAAVTKAIKSLVKQDMLAGTKDTVDARVTYFELTELAKPIASEHTHHHDETLNVYN 120

Query: 121 RLVNHFSKDEKVVLERFLDLFSRELEG 147
           RL+  FS  E  ++++F+ +F+ ELEG
Sbjct: 121 RLLQKFSAKELEIVDKFVTVFAEELEG 147
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 107

A DNA sequence (GBSx0110) was identified in *S.agalactiae* <SEQ ID 355> which encodes the amino acid sequence <SEQ ID 356>. This protein is predicted to be AdcC protein. Analysis of this protein sequence reveals the following:

```
Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1089(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA96186 GB:Z71552 AdcC protein [Streptococcus pneumoniae]
 Identities = 182/231 (78%), Positives = 206/231 (88%)

Query: 1   MRYITVSGLTFQYDSDPVLEGVNYHLDSGEFVTLTGENGAAKSTLIKATLGILTPKVGTV 60
           MRYITV  L+F YD +PVLE +NY +DSGEFVTLTGENGAAK+TLIKA+LGIL P++G V
Sbjct: 1   MRYITVEDLSFYYDKEPVLEHINYCVDSGEFVTLTGENGAAKTTLIKASLGILQPRIGKV 60

Query: 61  NISKENKEGKKLRIAYLPQQIASFNAGFPSSVYEFVKSGRYPRNGWFRRLTKHDEEHIRV 120
            ISK N +GKKLRIAYLPQQIASFNAGFPS+VYEFVKSGRYPR GWFRRL  HDEEHI+
Sbjct: 61  AISKTNTQGKKLRIAYLPQQIASFNAGFPSTVYEFVKSGRYPRKGWFRRLNAHDEEHIKA 120

Query: 121 SLEAVGMWDNRHKKIGSLSGGQKQRAVIARMFASDPDIFVLDEPTTGMDAGTTEKFYELM 180
           SL++VGMW++R K++GSLSGGQKQRAVIARMFASDPD+F+LDEPTTGMDAG+  +FYELM
Sbjct: 121 SLDSVGMWEHRDKRLGSLSGGQKQRAVIARMFASDPDVFILDEPTTGMDAGSKNEFYELM 180

Query: 181 HHNAHKHGKSVLMITHDPDEVKGYADRNIHLVRNQSLPWRCFNVHTNEMEV 231
           HH+AH HGK+VLMITHDP+EVK YADRNIHLVRNQ  PWRCFNVH N  EV
Sbjct: 181 HHSAHHHGKAVLMITHDPEEVKDYADRNIHLVRNQDSPWRCFNVHENGQEV 231
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 357> which encodes the amino acid sequence <SEQ ID 358>. Analysis of this protein sequence reveals the following:

```
Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2722(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 190/232 (81%), Positives = 214/232 (91%)

Query: 1   MRYITVSGLTFQYDSDPVLEGVNYHLDSGEFVTLTGENGAAKSTLIKATLGILTPKVGTV 60
           MRYI+V  L+FQY+S+PVLEG+ YHLDSGEFVT+TGENGAAKSTLIKATLGIL PK G V
Sbjct: 1   MRYISVKNLSFQYESEPVLEGITYHLDSGEFVTMTGENGAAKSTLIKATLGILQPKAGRV 60

Query: 61  NISKENKEGKKLRIAYLPQQIASFNAGFPSSVYEFVKSGRYPRNGWFRRLTKHDEEHIRV 120
            I+K+NK+GK+LRIAYLPQQ+ASFNAGFPS+VYEFVKSGRYPR+GWFR L KHDEEH++
Sbjct: 61  TIAKKNKDGKQLRIAYLPQQVASFNAGFPSTVYEFVKSGRYPRSGWFRHLNKHDEEHVQA 120

Query: 121 SLEAVGMWDNRHKKIGSLSGGQKQRAVIARMFASDPDIFVLDEPTTGMDAGTTEKFYELM 180
```

```
            SLEAVGMW+NRHK+IGSLSGGQKQR VIARMFASDPDIFVLDEPTTGMD+GTT+ FYELM
Sbjct: 121 SLEAVGMWENRHKRIGSLSGGQKQRVVIARMFASDPDIFVLDEPTTGMDSGTTDTFYELM 180

Query: 181 HHNAHKHGKSVLMITHDPDEVKGYADRNIHLVRNQSLPWRCFNVHTNEMEVE 232
           HH+AH+HGKSVLMITHDP+EVK YADRNIHLVRNQ LPWRCFN+H  E + E
Sbjct: 181 HHSAHQHGKSVLMITHDPEEVKAYADRNIHLVRNQKLPWRCFNIHEAETDDE 232
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 108

A DNA sequence (GBSx0111) was identified in *S.agalactiae* <SEQ ID 359> which encodes the amino acid sequence <SEQ ID 360>. Analysis of this protein sequence reveals the following:

```
Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2299(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 109

A DNA sequence (GBSx0112) was identified in *S.agalactiae* <SEQ ID 361> which encodes the amino acid sequence <SEQ ID 362>. This protein is predicted to be AdcB protein (znuB). Analysis of this protein sequence reveals the following:

```
Possible site: 36

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-14.33   Transmembrane  145 - 161 ( 136 - 172)
   INTEGRAL    Likelihood =-11.57   Transmembrane   29 -  45 (  20 -  47)
   INTEGRAL    Likelihood =-10.56   Transmembrane  261 - 277 ( 255 - 280)
   INTEGRAL    Likelihood = -8.70   Transmembrane  231 - 247 ( 227 - 253)
   INTEGRAL    Likelihood = -5.63   Transmembrane  101 - 117 (  99 - 121)
   INTEGRAL    Likelihood = -4.94   Transmembrane  186 - 202 ( 183 - 225)
   INTEGRAL    Likelihood = -3.82   Transmembrane   55 -  71 (  54 -  74)
   INTEGRAL    Likelihood = -3.61   Transmembrane  206 - 222 ( 203 - 225)
   INTEGRAL    Likelihood = -3.03   Transmembrane   78 -  94 (  75 -  94)

----- Final Results -----
               bacterial membrane --- Certainty=0.6731(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9487> which encodes amino acid sequence <SEQ ID 9488> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA96187 GB:Z71552 AdcB protein [Streptococcus pneumoniae]
 Identities = 197/263 (74%), Positives = 236/263 (88%)
```

```
Query: 13  LLDMLSYDFMQRALLAVVAISIFAPILGIFLILRRQSLMSDTLSHVSLAGVALGVVLGIS 72
           +L +LSYDF+QRA LAV+A+S+F+P+LG FLILRRQSLMSDTLSHVSL+GVA G+VLGIS
Sbjct: 1   MLSLLSYDFIQRAFLAVIAMSLFSPVLGTFLILRRQSLMSDTLSHVSLSGVAFGLVLGIS 60

Query: 73  PTWSTIFVVTLAAVVLEYLRTVYKHYMEISTAILMSMGLAISLIVMSKAHNVGNVSLEQY 132
           PT STI +V +AAV LEYLRTVYK +MEI TAILMS GLA+SLIVMSK  +  ++SL+QY
Sbjct: 61  PTVSTIAIVLIAAVFLEYLRTVYKSFMEIGTAILMSTGLAVSLIVMSKGKSSSSMSLDQY 120

Query: 133 LFGSIITIGKEQVIALFVIALITFILTILFIRPMYILTFDEDTAFVDGLPVRTMSILFNV 192
           LFGSI+TI +EQVI+LFVIA +  ILT LF+RPMYILTFDEDTAFVDGLPVRTMSILFN+
Sbjct: 121 LFGSIVTISEEQVISLFVIAAVVLILTFLFLRPMYILTFDEDTAFVDGLPVRTMSILFNM 180

Query: 193 VTGIAIALTIPAAGALLVSTIMVLPASIAMRLGRNFKTVIFLGMLIGFVGMVAGIFLSYY 252
           VTG+AIAL IPAAGALLVSTIMVLPASIA+RLG+NFK+V+ L   IGF+GMVAG+++SYY
Sbjct: 181 VTGVAIALMIPAAGALLVSTIMVLPASIALRLGKNFKSVMLLASAIGFLGMVAGLYISYY 240

Query: 253 WETPASATITMIFIGIFLLVSLV 275
            ETPASA+IT+IF+ +F+L+SLV
Sbjct: 241 AETPASASITIIFVTVFILISLV 263
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 363> which encodes the amino acid sequence <SEQ ID 364>. Analysis of this protein sequence reveals the following:

```
     Possible site: 18
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-14.97   Transmembrane  135 - 151 ( 123 - 162)
   INTEGRAL    Likelihood = -9.08   Transmembrane   68 -  84 (  44 -  86)
   INTEGRAL    Likelihood = -6.95   Transmembrane   20 -  36 (  19 -  37)
   INTEGRAL    Likelihood = -6.90   Transmembrane  251 - 267 ( 245 - 270)
   INTEGRAL    Likelihood = -6.58   Transmembrane  221 - 237 ( 217 - 243)
   INTEGRAL    Likelihood = -6.42   Transmembrane   91 - 107 (  89 - 111)
   INTEGRAL    Likelihood = -4.78   Transmembrane  176 - 192 ( 171 - 215)
   INTEGRAL    Likelihood = -3.82   Transmembrane   45 -  61 (  44 -  67)
   INTEGRAL    Likelihood = -3.61   Transmembrane  196 - 212 ( 193 - 215)

----- Final Results -----
               bacterial membrane --- Certainty=0.6986(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA96187 GB:Z71552 AdcB protein [Streptococcus pneumoniae]
 Identities = 195/262 (74%), Positives = 239/262 (90%)

Query: 3   MLDILFYDFMQRAVMAVVAISIFAPILGIFLILRRQSLMSDTLSHVSLAGVALGVVLGIS 62
           ML +L YDF+QRA +AV+A+S+F+P+LG FLILRRQSLMSDTLSHVSL+GVA G+VLGIS
Sbjct: 1   MLSLLSYDFIQRAFLAVIAMSLFSPVLGTFLILRRQSLMSDTLSHVSLSGVAFGLVLGIS 60

Query: 63  PTITTIIVVVLAAILLEYLRVVYKHYMEISTAILMSLGLALSLIIMSKSHSSSSMSLEQY 122
           PT++TI +V++AA+ LEYLR VYK +MEI TAILMS GLA+SLI+MSK  SSSSMSL+QY
Sbjct: 61  PTVSTIAIVLIAAVFLEYLRTVYKSFMEIGTAILMSTGLAVSLIVMSKGKSSSSMSLDQY 120

Query: 123 LFGSIITISMEQVVALFAIAAIILILTVLFIRPMYILTFDEDTAFVDGLPVRLMSVLFNI 182
           LFGSI+TIS EQV++LF IAA++LILT LF+RPMYILTFDEDTAFVDGLPVR MS+LFN+
Sbjct: 121 LFGSIVTISEEQVISLFVIAAVVLILTFLFLRPMYILTFDEDTAFVDGLPVRTMSILFNM 180

Query: 183 VTGVAIALTIPAAGALLVSTIMVLPASIAMRLGKNFKTVILLGIVIGFSGMLSGIFLSYF 242
           VTGVAIAL IPAAGALLVSTIMVLPASIA+RLGKNFK+V+LL   IGF GM++G+++SY+
Sbjct: 181 VTGVAIALMIPAAGALLVSTIMVLPASIALRLGKNFKSVMLLASAIGFLGMVAGLYISYY 240

Query: 243 FETPASATITMIFISIFLLVSL 264
            ETPASA+IT+IF+++F+L+SL
Sbjct: 241 AETPASASITIIFVTVFILISL 262
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 223/270 (82%), Positives = 252/270 (92%)

Query: 12  MLLDMLSYDFMQRALLAVVAISIFAPILGIFLILRRQSLMSDTLSHVSLAGVALGVVLGI 71
           ++LD+L YDFMQRA++AVVAISIFAPILGIFLILRRQSLMSDTLSHVSLAGVALGVVLGI
Sbjct: 2   VMLDILFYDFMQRAVMAVVAISIFAPILGIFLILRRQSLMSDTLSHVSLAGVALGVVLGI 61

Query: 72  SPTWSTIFVVTLAAVVLEYLRTVYKHYMEISTAILMSMGLAISLIVMSKAHNVGNVSLEQ 131
           SPT +TI VV LAA++LEYLR VYKHYMEISTAILMS+GLA+SLI+MSK+H+  ++SLEQ
Sbjct: 62  SPTITTIIVVVLAAILLEYLRVVYKHYMEISTAILMSLGLALSLIIMSKSHSSSSMSLEQ 121

Query: 132 YLFGSIITIGKEQVIALFVIALITFILTILFIRPMYILTFDEDTAFVDGLPVRTMSILFN 191
           YLFGSIITI  EQV+ALF IA I  ILT+LFIRPMYILTFDEDTAFVDGLPVR MS+LFN
Sbjct: 122 YLFGSIITISMEQVVALFAIAAIILILTVLFIRPMYILTFDEDTAFVDGLPVRLMSVLFN 181

Query: 192 VVTGIAIALTIPAAGALLVSTIMVLPASIAMRLGRNFKTVIFLGMLIGFVGMVAGIFLSY 251
           +VTG+AIALTIPAAGALLVSTIMVLPASIAMRLG+NFKTVI LG++IGF GM++GIFLSY
Sbjct: 182 IVTGVAIALTIPAAGALLVSTIMVLPASIAMRLGKNFKTVILLGIVIGFSGMLSGIFLSY 241

Query: 252 YWETPASATITMIFIGIFLLVSLVGLLRKR 281
           ++ETPASATITMIFI IFLLVSL G+L+KR
Sbjct: 242 FFETPASATITMIFISIFLLVSLGGMLKKR 271
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 110

A DNA sequence (GBSx0113) was identified in *S.agalactiae* <SEQ ID 365> which encodes the amino acid sequence <SEQ ID 366>. This protein is predicted to be streptodornase. Analysis of this protein sequence reveals the following:

```
Possible site: 59

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2601(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA59264 GB:X84793 streptodornase [Streptococcus pyogenes]
 Identities = 58/167 (34%), Positives = 85/167 (50%), Gaps = 30/167 (17%)

Query: 2   TPIYEGNNLVPSRVELQYVGIDKQGKLLEIKLGGGKEQVDEYGVTTVTLENTSPLAKIDY 61
           TP+Y+G+ L+P  V +  +  D                +DE    TV + N      IDY
Sbjct: 245 TPVYQGSELLPRAVLVSALSSDGF--------------IDE----TVRVFNNVAGFNIDY 286

Query: 62  KTGMLIKEDGKQAEEGEDPNSDADENEAAIE-SASDIEENTNTNTSESDTNNVAPQNRIV 120
           + G L+ E          P ++ D  E  +E +   IE+  +T+T + D  N++ Q + V
Sbjct: 287 QNGGLLTES---------PVTETDNVEENVEDNIETIEDEVDTDTLKKDDENISLQ-KTV 336

Query: 121 YVANKGRSNTYWYSLENI-KNANTANIVQMTEQEALNQHKHHSTTEA 166
           YVA+ G SN YWYS EN+ KN N   +V+M+EQ AL + KHHS  EA
Sbjct: 337 YVASSGLSNVYWYSKENMPKNVNLDKVVEMSEQTALARGKHHSAQEA 383
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 367> which encodes the amino acid sequence <SEQ ID 368>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have a cleavable N-term signal seq.
```

```
----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 51/90 (56%), Positives = 66/90 (72%), Gaps = 4/90 (4%)

Query: 1   MTPIYEGNNLVPSRVELQYVGIDKQGKLLEIKLGGGKEQVDEYGVTTVTLENTSPLAKID 60
           +TP+Y  N LVP +V LQYVGID+ G LL+IKLG  KE VD +GVT+VTL+N SPLA++D
Sbjct: 182 VTPVYHKNELVPRQVVLQYVGIDENGDLLQIKLGSEKESVDNFGVTSVTLDNVSPLAELD 241

Query: 61  YKTGMLIKEDGKQAEEGEDPNSDADENEAA 90
           Y+TGM++  D  Q E  ED N + +E E A
Sbjct: 242 YQTGMML--DSTQNE--EDSNLETEEFEEA 267
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 111

A DNA sequence (GBSx0114) was identified in *S.agalactiae* <SEQ ID 369> which encodes the amino acid sequence <SEQ ID 370>. This protein is predicted to be tyrosyl-tRNA synthetase (tyrS-1). Analysis of this protein sequence reveals the following:

```
Possible site: 60

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3618(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC00303 GB:AF008220 tyrosine tRNA synthetase [Bacillus subtilis]
 Identities = 234/420 (55%), Positives = 311/420 (73%), Gaps = 2/420 (0%)

Query: 2   NIFDELKERGLVFQTTDEDALRKALEEGSVSYYTGYDPTADSLHLGHLVAILTSRRLQLA 61
           N+ ++L  RGL+ Q TDE+ L K L E  +  Y+G+DPTADSLH+GHL+ ILT RR QLA
Sbjct: 3   NLLEDLSFRGLIQQMTDEEGLNKQLNEEKIRLYSGFDPTADSLHIGHLLPILTLRRFQLA 62

Query: 62  GHKPYALVGGATGLIGDPSFKDVERSLQTKKTVVSWGNKIRGQLSNFLEFETGDNKAVLV 121
           GH P ALVGGATGLIGDPS K  ER+L T   V  W  KI+ QLS FL+FE  +N AV+
Sbjct: 63  GHHPIALVGGATGLIGDPSGKKAERTLNTADIVSEWSQKIKNQLSRFLDFEAAENPAVIA 122

Query: 122 NNYDWFSNISFIDFLRDVGKYFTVNYMMSKESVKKRIETGISYTEFAYQIMQGYDFYELN 181
           NN+DW   ++ IDFLRDVGK F +NYM++K++V  RIE+GISYTEF+Y I+Q YDF  L
Sbjct: 123 NNFDWIGKMNVIDFLRDVGKNFGINYMLAKDTVSSRIESGISYTEFSYMILQSYDFLNLY 182

Query: 182 KNYNVTLQIGGSDQWGNMTAGTELIRR--KSNGVSHVMTVPLITDSTGKKFGKSEGNAVW 239
           ++ N  LQIGGSDQWGN+TAG ELIR+  +     +  +T+PL+T + G KFGK+EG A+W
Sbjct: 183 RDKNCKLQIGGSDQWGNITAGLELIRKSEEEGAKAFGLTIPLVTKADGTKFGKTEGGAIW 242

Query: 240 LDADKTSPYEMYQFWLNVMDADAVRFLKIFTFLSLKEIEDIRIQFEEAPHQRLAQKTLAR 299
           LD +KTSPYE YQFW+N  D D V++LK FTFLS +EIE    + E AP +R AQK LA
Sbjct: 243 LDKEKTSPYEFYQFWINTDDRDVVKYLKYFTFLSKEEIEAYAEKTETAPEKREAQKRLAE 302

Query: 300 EVVTLVHGEKAYKEAVNITEQLFAGNIKGLSVKELKQGLRGVPNYHVQTEDNLNIIDLLV 359
           EV +LVHG +A ++A+NI++ LF+GNIK LS +++K G + VP+  V +   L+++D+LV
Sbjct: 303 EVTSLVHGREALEQAINISQALFSGNIKELSAQDVKVGFKDVPSMEVDSTQELSLVDVLV 362

Query: 360 TSGVVNSKRQAREDVSNGAIYINGDRIQDLEYTISENDKLENEITVIRRGKKKYFVLNFK 419
```

```
            S +  SKRQARED+ NGA+YING+R  ++ YT+S  D++EN+ TV+RRGKKKYF++ +K
Sbjct: 363 QSKLSPSKRQAREDIQNGAVYINGERQTEINYTLSGEDRIENQFTVLRRGKKKYFLVTYK 422
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 371> which encodes the amino acid sequence <SEQ ID 372>. Analysis of this protein sequence reveals the following:

```
Possible site: 37

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2340(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 344/418 (82%), Positives = 377/418 (89%)

Query: 1   MNIFDELKERGLVFQTTDEDALRKALEEGSVSYYTGYDPTADSLHLGHLVAILTSRRLQL 60
           MNIF+ELK RGLVFQTTDE AL KAL EG VSYYTGYDPTADSLHLGHLVAILTSRRLQL
Sbjct: 1   MNIFEELKARGLVFQTTDEQALVKALTEGQVSYYTGYDPTADSLHLGHLVAILTSRRLQL 60

Query: 61  AGHKPYALVGGATGLIGDPSFKDVERSLQTKKTVVSWGNKIRGQLSNFLEFETGDNKAVL 120
           AGHKPYALVGGATGLIGDPSFKD ERSLQTK+TV+ W +KI+GQLS FL+FE GDNKA L
Sbjct: 61  AGHKPYALVGGATGLIGDPSFKDAERSLQTKETVLEWSDKIKGQLSTFLDFENGDNKAEL 120

Query: 121 VNNYDWFSNISFIDFLRDVGKYFTVNYMMSKESVKKRIETGISYTEFAYQIMQGYDFYEL 180
           VNNYDWFS ISFIDFLRDVGKYFTVNYMMSK+SVKKRIETGISYTEFAYQIMQGYDFYEL
Sbjct: 121 VNNYDWFSQISFIDFLRDVGKYFTVNYMMSKDSVKKRIETGISYTEFAYQIMQGYDFYEL 180

Query: 181 NKNYNVTLQIGGSDQWGNMTAGTELIRRKSNGVSHVMTVPLITDSTGKKFGKSEGNAVWL 240
           N  +NVTLQIGGSDQWGNMTAGTEL+R+K++   HVMTVPLITDSTGKKFGKSEGNAVWL
Sbjct: 181 NDKHNVTLQIGGSDQWGNMTAGTELLRKKADKTGHVMTVPLITDSTGKKFGKSEGNAVWL 240

Query: 241 DADKTSPYEMYQFWLNVMDADAVRFLKIFTFLSLKEIEDIRIQFEEAPHQRLAQKTLARE 300
           DADKTSPYEMYQFWLNVMD DAVRFLKIFTFLSL EI +I  QF  A H+RLAQKTLARE
Sbjct: 241 DADKTSPYEMYQFWLNVMDDDAVRFLKIFTFLSLDEIAEIETQFNAARHERLAQKTLARE 300

Query: 301 VVTLVHGEKAYKEAVNITEQLFAGNIKGLSVKELKQGLRGVPNYHVQTEDNLNIIDLLVT 360
           VVTLVHGE+AYK+A+NITEQLFAGNIK LS  ELKQGL  VPNYHVQ+ DN NI+++LV
Sbjct: 301 VVTLVHGEEAYKQALNITEQLFAGNIKNLSANELKQGLSNVPNYHVQSIDNHNIVEILVA 360

Query: 361 SGVVNSKRQAREDVSNGAIYINGDRIQDLEYTISENDKLENEITVIRRGKKKYFVLNF 418
           + +  SKRQAREDV NGAIYINGDR+QDL+Y +S +DK+++++TVIRRGKKKY VL +
Sbjct: 361 AKISPSKRQAREDVQNGAIYINGDRVQDLDYQLSNDDKIDDQLTVIRRGKKKYAVLTY 418
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 112

A DNA sequence (GBSx0115) was identified in *S.agalactiae* <SEQ ID 373> which encodes the amino acid sequence <SEQ ID 374>. Analysis of this protein sequence reveals the following:

```
Possible site: 53

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-12.21   Transmembrane   36 -  52 (  23 -  59)

----- Final Results -----
              bacterial membrane --- Certainty=0.5883(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF04736 GB:AF101781 penicillin-binding protein 1b
           [Streptococcus pneumoniae]
 Identities = 445/769 (57%), Positives = 581/769 (74%), Gaps = 9/769 (1%)

Query: 3   KGNKKLNSSKLGDYTP----LEFGSIFLRI---VKLLSDFIYVIILLFVMLGVGLAVGYL 55
           K  K     K G  T     L+  +IF  I   +K L + ++V+  L  MLG G+A+GY
Sbjct: 21  KNKKSARPGKKGSSTKKSKTLDKSAIFPAILLSIKALFNLLFVLGFLGGMLGAGIALGYG 80

Query: 56  ASQVDSVKVPSKNSLVTQVNTLTRVSRLTYSDKSQISEIATDLQRTPVAKDAISDNIKKA 115
            +  D V+VP     LV QV  ++ +S +TYSD + I+ I +DL RT ++ + IS+N+KKA
Sbjct: 81  VALFDKVRVPQTEELVNQVKDISSISEITYSDGTVIASIESDLLRTSISSEQISENLKKA 140

Query: 116 IIATEDENFNDHKGVVPKAVLRAAAGSVLGFGESSGGSTLTQQLLKQQILGDDPSFKRKS 175
           IIATEDE+F +HKGVVPKAV+RA  G  +G G SSGGSTLTQQL+KQQ++GD P+  RK+
Sbjct: 141 IIATEDEHFKEHKGVVPKAVIRATLGKFVGLGSSSGGSTLTQQLIKQQVVGDAPTLARKA 200

Query: 176 KEIIYALALERYMDKDSILSDYLNVSPFGRNNKGQNIAGIEEAAQGIFGVSAKDLTIPQA 235
            EI+ ALALER M+KD IL+ YLNV+PFGRNNKGQNIAG  +AA+GIFGV A  LT+PQA
Sbjct: 201 AEIVDALALERAMNKDEILTTYLNVAPFGRNNKGQNIAGARQAAEGIFGVDASQLTVPQA 260

Query: 236 AFLAGLPQSPIVYSPYTADAQLKSDKDLSFGIKRQKNVLYNMYRTRALTKDEYKSYKDYD 295
           AFLAGLPQSPI YSPY    +LKSD+DL  G++R K VLY+MYRT AL+KDEY  YKDYD
Sbjct: 261 AFLAGLPQSPITYSPYENTGELKSDEDLEIGLRRAKAVLYSMYRTGALSKDEYSQYKDYD 320

Query: 296 IKKDFIKPAVATTNHHDYLYYSALSEAQKVMYNYLIKKDNVSEHDLKNDETRATYRHRAI 355
           +K+DF+     T    DYLY++ L+EAQ+ MY+YL ++DNVS  +LKN+ T+  YR  A
Sbjct: 321 LKQDFLPSGTVTGISRDYLYFTTLAEAQERMYDYLAQRDNVSAKELKNEATQKFYRDLAA 380

Query: 356 EEIQQGGYTIKTTINKSVYQAMQDAAAQYGGLLDDGTGKVQMGNVLTDNSSGAIIGFIGG 415
           +EI+ GGY I TTI++ ++ AMQ A A YG LLDDGTG+V++GNVL DN +GAI+GF+GG
Sbjct: 381 KEIENGGYKITTTIDQKIHSAMQSAVADYGYLLDDGTGRVEVGNVLMDNQTGAILGFVGG 440

Query: 416 RNYSENQNNHAFDTARSPGSSIKPILPYGIAIDQGMLGSGSVLSNYPTTYSSGEKIMHAD 475
           RNY ENQNNHAFDT RSP S+ KP+L YGIAIDQG++GS ++LSNYPT +++G  IM+A+
Sbjct: 441 RNYQENQNNHAFDTKRSPASTTKPLLAYGIAIDQGLMGSETILSNYPTNFANGNPIMYAN 500

Query: 476 EEGTAMVNLQESLDISWNIPAFWTYKMLRDRGVDVKNYMEKLDYPIENFGIESLPLGGGI 535
            +GT M+ L E+L+ SWNIPA+WTY+MLR+ GVDVK YMEK+ Y I  +GIESLP+GGGI
Sbjct: 501 SKGTGMMTLGEALNYSWNIPAYWTYRMLRENGVDVKGYMEKMGYEIPEYGIESLPMGGGI 560

Query: 536 DTSVAQQTNLYQMIANGGVYHKQYMIESIEDSNGKVIYNHESKPVRVFSKATATILQQLL 595
           + +VAQ TN YQ +AN GVYH++++I  IE ++G+V+Y ++ KPV+V+SKATATI+Q LL
Sbjct: 561 EVTVAQHTNGYQTLANNGVYHQKHVISKIEAADGRVVYEYQDKPVQVYSKATATIMQGLL 620

Query: 596 HGPINSGKTTTFKNRLQGLNSGLAGVDWIGKTGTTNSTSDVWLMLSTPKVTLGGWAGHDN 655
              ++S  TTTFK+ L  LN  LA  DWIGKTGTTN   ++WLMLSTP++TLGGW GHD+
Sbjct: 621 REVLSSRVTTTFKSNLTSLNPTLANADWIGKTGTTNQDENMWLMLSTPRLTLGGWIGHDD 680

Query: 656 NASLAKLTGYNNNNANYMAHLVNAINNADGNTFGKSERFRLDDSVIKAKVLKSTGLQPGVV 715
           N SL++  GY+NN+NYMAHLVNAI  A  + +G +ERF LD SV+K++VLKSTG +PG V
Sbjct: 681 NHSLSRRAGYSNNSNYMAHLVNAIQQASPSIWG-NERFALDPSVVKSEVLKSTGQKPGKV 739

Query: 716 TVNGRRITVGGESTTSYWA-KNGPGTMTYRFAIGGTDSDYQKAWSTLGG 763
           +V G+ + V G + TSYWA K+G     +YRFAIGG+D+DYQ AWS++ G
Sbjct: 740 SVEGKEVEVTGSTVTSYWANKSGAPATSYRFAIGGSDADYQNAWSSIVG 788
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 375> which encodes the amino acid sequence <SEQ ID 376>. Analysis of this protein sequence reveals the following:

```
      Possible site: 57

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.83   Transmembrane   39 -  55 ( 32 -  60)

----- Final Results -----
```

```
bacterial membrane --- Certainty=0.2932(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF04736 GB:AF101781 penicillin-binding protein 1b
          [Streptococcus pneumoniae]
 Identities = 438/739 (59%), Positives = 580/739 (78%), Gaps = 2/739 (0%)

Query: 27  PVLLRTLRLLSNFFYIVIFLFGMMGFGMAFGYLASQIESVKVPSKESLVKQVESLTMISQ 86
           P +L +++ L N  +++ FL GM+G G+A GY  +  + V+VP  E LV QV+ ++ IS+
Sbjct: 48  PAILLSIKALFNLLFVLGFLGGMLGAGIALGYGVALFDKVRVPQTEELVNQVKDISSISE 107

Query: 87  MNYSDNSLISTLDTDLLRTPVANDAISENIKKAIVSTEDEHFQEHKGIVPKAVFRATLAS 146
           + YSD ++I+++++DLLRT ++++ ISEN+KKAI++TEDEHF+EHKG+VPKAV RATL
Sbjct: 108 ITYSDGTVIASIESDLLRTSISSEQISENLKKAIIATEDEHFKEHKGVVPKAVIRATLGK 167

Query: 147 VLGFGEASGGSTLTQQLVKQQVLGDDPTFKRKSKEIVYALALERYMSKDNILCDYLNVSP 206
            +G G +SGGSTLTQQL+KQQV+GD PT  RK+ EIV ALALER M+KD IL  YLNV+P
Sbjct: 168 FVGLGSSSGGSTLTQQLIKQQVVGDAPTLARKAAEIVDALALERAMNKDEILTTYLNVAP 227

Query: 207 FGRNNKGQNIAGVEEAARGIFGVSAKDLTVPQAAFLAGLPQSPIVYSPYLSTGQLKSEKD 266
           FGRNNKGQNIAG  +AA GIFGV A  LTVPQAAFLAGLPQSPI YSPY +TG+LKS++D
Sbjct: 228 FGRNNKGQNIAGARQAAEGIFGVDASQLTVPQAAFLAGLPQSPITYSPYENTGELKSDED 287

Query: 267 MAYGIKRQQNVLFNMYRTGVLSKKEYEDYKAYPIQKDFIQPGSAIVNNHDYLYYTVLADA 326
           +  G++R + VL++MYRTG LSK EY  YK Y +++DF+  G+    + DYLY+T LA+A
Sbjct: 288 LEIGLRRAKAVLYSMYRTGALSKDEYSQYKDYDLKQDFLPSGTVTGISRDYLYFTTLAEA 347

Query: 327 KKAMYSYLIKRDKVSSRDLKNDETKAAYEERALTELQQGGYTITTTINKPIYNAMQTAAA 386
           ++ MY YL +RD VS+++LKN+ T+  Y + A  E++ GGY ITTTI++ I++AMQ+A A
Sbjct: 348 QERMYDYLAQRDNVSAKELKNEATQKFYRDLAAKEIENGGYKITTTIDQKIHSAMQSAVA 407

Query: 387 QFGGLLDDGTGTVQMGNVLTDNATGAVLGFVGGRDYALNQNNHAFNTVRSPGSSIKPIIA 446
            +G LLDDGTG V++GNVL DN TGA+LGFVGGR+Y  NQNNHAF+T RSP S+ KP++A
Sbjct: 408 DYGYLLDDGTGRVEVGNVLMDNQTGAILGFVGGRNYQENQNNHAFDTKRSPASTTKPLLA 467

Query: 447 YGPAIDQGLMGSASVLSNYPTTYSSGQKIMHADSEGTAMMPLQEALNTSWNIPAFWTQKL 506
           YG AIDQGLMGS ++LSNYPT +++G  IM+A+S+GT MM L EALN SWNIPA+WT ++
Sbjct: 468 YGIAIDQGLMGSETILSNYPTNFANGNPIMYANSKGTGMMTLGEALNYSWNIPAYWTYRM 527

Query: 507 LREKGVDVENYMTKMGYKIADYSIESLPLGGGIEVSVAQQTNAYQMLSNNGLYQKQYIVD 566
           LRE GVDV+ YM KMGY+I +Y IESLP+GGGIEV+VAQ TN YQ L+NNG+Y +++++
Sbjct: 528 LRENGVDVKGYMEKMGYEIPEYGIESLPMGGGIEVTVAQHTNGYQTLANNGVYHQKHVIS 587

Query: 567 KITASDGTVVYKHENKPIRIFSAATATILQELLRGPITSGATTTFKNRLAAINPWLANAD 626
           KI A+DG VVY++++KP+++++S ATATI+Q LLR  ++S  TTTFK+ L ++NP LANAD
Sbjct: 588 KIEAADGRVVYEYQDKPVQVYSKATATIMQGLLREVLSSRVTTTFKSNLTSLNPTLANAD 647

Query: 627 WIGKTGTTENYTDVWLVLSTPKVTLGGWAGHDDNTSLAPLTGYNNNSNYLAYLANAINQA 686

           WIGKTGTT    ++WL+LSTP++TLGGW GHDDN SL+   GY+NNSNY+A+L NAI QA
Sbjct: 648 WIGKTGTTNQDENMWLMLSTPRLTLGGWIGHDDNHSLSRRAGYSNNSNYMAHLVNAIQQA 707

Query: 687 DPNVIGVGQRFNLDPGVIKANVLKSTGLQPGTVNVNGHTFSVGGEMTTSLWSQK-GPGAM 745
            P++ G  +RF LDP V+K+ VLKSTG +PG V+V G    V G   TS W+ K G  A
Sbjct: 708 SPSIWG-NERFALDPSVVKSEVLKSTGQKPGKVSVEGKEVEVTGSTVTSYWANKSGAPAT 766

Query: 746 TYRFAIGGTDADYQKAWGN 764
           +YRFAIGG+DADYQ AW +
Sbjct: 767 SYRFAIGGSDADYQNAWSS 785
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 531/760 (69%), Positives = 639/760 (83%), Gaps = 3/760 (0%)

Query: 6   KKLNSSKLGDYTPLEFGSIFLRIVKLLSDFIYVIILLFVMLGVGLAVGYLASQVDSVKVP 65
```

```
            K+++  +LG    L+ G + LR ++LLS+F Y++I LF M+G G+A GYLASQ++SVKVP
Sbjct: 13   KRISHQRLG---LLDLGPVLLRTLRLLSNFFYIVIFLFGMMGFGMAFGYLASQIESVKVP 69

Query: 66   SKNSLVTQVNTLTRVSRLTYSDKSQISEIATDLQRTPVAKDAISDNIKKAIIATEDENFN 125
            SK SLV QV +LT +S++ YSD S IS + TDL RTPVA DAIS+NIKKAI++TEDE+F
Sbjct: 70   SKESLVKQVESLTMISQMNYSDNSLISTLDTDLLRTPVANDAISENIKKAIVSTEDEHFQ 129

Query: 126  DHKGVVPKAVLRAAAGSVLGFGESSGGSTLTQQLLKQQILGDDPSFKRKSKEIIYALALE 185
            +HKG+VPKAV RA   SVLGFGE+SGGSTLTQQL+KQQ+LGDDP+FKRKSKEI+YALALE
Sbjct: 130  EHKGIVPKAVFRATLASVLGFGEASGGSTLTQQLVKQQVLGDDPTFKRKSKEIVYALALE 189

Query: 186  RYMDKDSILSDYLNVSPFGRNNKGQNIAGIEEAAQGIFGVSAKDLTIPQAAFLAGLPQSP 245
            RYM KD+IL DYLNVSPFGRNNKGQNIAG+EEAA+GIFGVSAKDLT+PQAAFLAGLPQSP
Sbjct: 190  RYMSKDNILCDYLNVSPFGRNNKGQNIAGVEEAARGIFGVSAKDLTVPQAAFLAGLPQSP 249

Query: 246  IVYSPYTADAQLKSDKDLSFGIKRQKNVLYNMYRTRALTKDEYKSYKDYDIKKDFIKPAV 305
            IVYSPY +  QLKS+KD+++GIKRQ+NVL+NMYRT  L+K EY+ YK Y I+KDFI+P
Sbjct: 250  IVYSPYLSTGQLKSEKDMAYGIKRQQNVLFNMYRTGVLSKKEYEDYKAYPIQKDFIQPGS 309

Query: 306  ATTNHHDYLYYSALSEAQKVMYNYLIKKDNVSEHDLKNDETRATYRHRAIEEIQQGGYTI 365
            A  N+HDYLYY+ L++A+K MY+YLIK+D VS  DLKNDET+A Y  RA+ E+QQGGYTI
Sbjct: 310  AIVNNHDYLYYTVLADAKKAMYSYLIKRDKVSSRDLKNDETKAAYEERALTELQQGGYTI 369

Query: 366  KTTINKSVYQAMQDAAAQYGGLLDDGTGKVQMGNVLTDNSSGAIIGFIGGRNYSENQNNH 425
             TTINK +Y AMQ AAAQ+GGLLDDGTG VQMGNVLTDN++GA++GF+GGR+Y+ NQNNH
Sbjct: 370  TTTINKPIYNAMQTAAAQFGGLLDDGTGTVQMGNVLTDNATGAVLGFVGGRDYALNQNNH 429

Query: 426  AFDTARSPGSSIKPILPYGIAIDQGMLGSGSVLSNYPTTYSSGEKIMHADEEGTAMVNLQ 485
            AF+T RSPGSSIKPI+ YG AIDQG++GS SVLSNYPTTYSSG+KIMHAD EGTAM+ LQ
Sbjct: 430  AFNTVRSPGSSIKPIIAYGPAIDQGLMGSASVLSNYPTTYSSGQKIMHADSEGTAMMPLQ 489

Query: 486  ESLDISWNIPAFWTYKMLRDRGVDVKNYMEKLDYPIENFGIESLPLGGGIDTSVAQQTNL 545
            E+L+ SWNIPAFWT K+LR++GVDV+NYM K+ Y I ++ IESLPLGGGI+ SVAQQTN
Sbjct: 490  EALNTSWNIPAFWTQKLLREKGVDVENYMTKMGYKIADYSIESLPLGGGIEVSVAQQTNA 549

Query: 546  YQMIANGGVYHKQYMIESIEDSNGKVIYNHESKPVRVFSKATATILQQLLHGPINSGKTT 605
            YQM++N G+Y KQY+++ I  S+G V+Y HE+KP+R+FS ATATILQ+LL GPI SG TT
Sbjct: 550  YQMLSNNGLYQKQYIVDKITASDGTVVYKHENKPIRIFSAATATILQELLRGPITSGATT 609

Query: 606  TFKNRLQGLNSGLAGVDWIGKTGTTNSTSDVWLMLSTPKVTLGGWAGHDNNASLAKLTGY 665
            TFKNRL  +N  LA  DWIGKTGTT + +DVWL+LSTPKVTLGGWAGHD+N SLA LTGY
Sbjct: 610  TFKNRLAAINPWLANADWIGKTGTTENYTDVWLVLSTPKVTLGGWAGHDDNTSLAPLTGY 669

Query: 666  NNNANYMAHLVNAINNADGNTFGKSERFRLDDSVIKAKVLKSTGLQPGVVTVNGRRITVG 725
            NNN+NY+A+L NAIN AD N  G  +RF LD  VIKA VLKSTGLQPG V VNG   +VG
Sbjct: 670  NNNSNYLAYLANAINQADPNVIGVGQRFNLDPGVIKANVLKSTGLQPGTVNVNGHTFSVG 729

Query: 726  GESTTSYWAKNGPGTMTYRFAIGGTDSDYQKAWSTLGGKR 765
            GE TTS W++ GPG MTYRFAIGGTD+DYQKAW   G ++
Sbjct: 730  GEMTTSLWSQKGPGAMTYRFAIGGTDADYQKAWGNFGFRK 769
```

SEQ ID 374 (GBS64d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 120 (lane 2-4; MW 107kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 120 (lane 5-7; MW 82kDa) and in Figure 179 (lane 2; MW 82kDa).

GBS64d-His was purified as shown in Figure 231, lane 7-8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 113

A DNA sequence (GBSx0116) was identified in *S.agalactiae* <SEQ ID 377> which encodes the amino acid sequence <SEQ ID 378>. This protein is predicted to be DNA-dependent RNA polymerase subunit beta (rpoB). Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3505(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB56706 GB:Y16468 DNA-dependent RNA polymerase subunit beta
            [Listeria monocytogenes]
 Identities = 814/1173 (69%), Positives = 978/1173 (82%), Gaps = 17/1173 (1%)

Query: 2    AGHEVQYGKHRTRRSFSRIKEVLDLPNLIEIQTDSFQDFLDAGLKEVFEDVLPISNFTDT 61
             +GH+V+YG+HRTRRSF+RI EVL+LPNLIEIQT S+Q FLD GL+E+F D+ PI +F
Sbjct: 5    SGHDVKYGRHRTRRSFARISEVLELPNLIEIQTASYQWFLDEGLREMFRDISPIEDFAGN 64

Query: 62   MDLEFVGYELKEPKYTLEEARIHDASYSAPIFVTFRLVNKETGEIKTQEVFFGDFPIMTE 121
             + LEF+ Y+L EPKY++EE++  DA+Y+AP+ V  RL+NKETGE+K QEVF GDFP+MTE
Sbjct: 65   LSLEFIDYDLGEPKYSVEESKNRDANYAAPLRVKLRLINKETGEVKDQEVFMGDFPLMTE 124

Query: 122  MGTFIINGGERIIVSQLVRSPGVYFNDKVDKNGKVGYGSTVIPNRGAWLELETDAKDIAY 181
             MGTFIING ER+IVSQLVRSPGVYFN K+DKNGK G+GSTVIPNRGAWLE ETDAKD+ +
Sbjct: 125  MGTFIINGAERVIVSQLVRSPGVYFNGKLDKNGKKGFGSTVIPNRGAWLEYETDAKDVVH 184

Query: 182  TRIDRTRKIPFTTLVRALGFSGDDEIVDIFGDSELVRNTIEKDIHKNPSDSRTDEALKEI 241
              RIDRTRK+P T L+RALGF  D EI+D+ GD++ +RNT+EKD   N      ++AL EI
Sbjct: 185  VRIDRTRKLPVTVLLRALGFGSDQEIIDLIGDNDYLRNTLEKDNTDN-----AEKALLEI 239

Query: 242  YERLRPGEPKTADSSRSLLVARFFDPRRYDLAAVGRYKINKKLNLKTRLLNQTIAENLVD 301
             YERLRPGEP T D++RSLLV+RFFDP+RYDLA+VGRYKINKKL+LK RL NQT+AE LVD
Sbjct: 240  YERLRPGEPPTVDNARSLLVSRFFDPKRYDLASVGRYKINKKLHLKNRLFNQTLAETLVD 299

Query: 302  GETGEILVEAGTVMTRDVIDSIAEHIDGDLNKFVYTPNDYAVVTEPVILQKFKVVAPTDP 361
              ETGEI+   G ++ R  +D I  +++  +      P D  V+ + V++Q  K+ AP D
Sbjct: 300  PETGEIIASKGDILDRRNLDQIIPNLENGVGFRTLRPTD-GVMEDSVLVQSIKIYAPNDE 358

Query: 362  DRVVTIVGNSNPEDKVRALTPADILAEMSYFLNLAEGIGKVDDIDHLGNRRIRAVGELLA 421
             ++ + I+GN+  E+ V+ +TP+DI++ +SYF NL  G+G  DDIDHLGNRR+R+VGELL
Sbjct: 359  EKEINIIGNAYIEENVKHITPSDIISSISYFFNLLHGVGDTDDIDHLGNRRLRSVGELLQ 418

Query: 422  NQFRIGLARMERNVRERMSVQDNEVLTPQQIINIRPVTAAVKEFFGSSQLSQFMDQHNPL 481
             NQFRIGL+RMER VRERMS+QD   +TPQQ+INIRPV A++KEFFGSSQLSQFMDQ NPL
Sbjct: 419  NQFRIGLSRMERVVRERMSIQDMTTITPQQLINIRPVVASIKEFFGSSQLSQFMDQTNPL 478

Query: 482  SELSHKRRLSALGPGGLTRDRAGYEVRDVHYTHYGRMCPIETPEGPNIGLINNLSSFGHL 541
              EL+HKRRLSALGPGGLTR+RAGYEVRDVHY+HYGRMCPIETPEGPNIGLIN+LSSF  +
Sbjct: 479  GELTHKRRLSALGPGGLTRERAGYEVRDVHYSHYGRMCPIETPEGPNIGLINSLSSFAKV 538

Query: 542  NKYGFIQTPYRKVDRSTGAVTNEIVWLTADEEDEFTVAQANSKLNEDGTFAEEIVMGRHQ 601
             NK+GFI+TPYR+VD  T  VT++I +LTADEED + VAQANSKL+E GTF EE VM R +
Sbjct: 539  NKFGFIETPYRRVDPETNRVTDKIDYLTADEEDNYVVAQANSKLDEQGTFTEEEVMARFR 598

Query: 602  GNNQEFPSSIVDFVDVSPKQVVAVATACIPFLENDDSNRALMGANMQRQAVPLIDPKAPY 661
               N       +D++DVSPKQVV+VATACIPFLENDDSNRALMGANMQRQAVPL+ P+AP+
Sbjct: 599  SENLAVEKERIDYMDVSPKQVVSVATACIPFLENDDSNRALMGANMQRQAVPLMHPEAPF 658

Query: 662  VGTGMEYQAAHDSGAAVIAKHDGRVIFSDAEKVEVRRED--------GSLDVYHVQKFRR 713
             VGTGME+ +A DSGAAV AKHDG V   +A ++ VRR          G +D Y ++KF R
Sbjct: 659  VGTGMEHVSAKDSGAAVTAKHDGIVEHVEAREIWVRRVSLVDGKEVTGGIDKYTLRKFVR 718
```

```
Query: 714  SNSGTAYNQRTLVKVGDLVEKGDFIADGPSMENGEMALGQNPVVAYMTWEGYNFEDAVIM 773
            SN GT YNQR  V  GD V KG+ + +GPSM++GE+ALG+N +VA+MTW+GYN+EDA+IM
Sbjct: 719  SNQGTCYNQRPNVAEGDRVVKGEILGNGPSMDSGELALGRNVLVAFMTWDGYNYEDAIIM 778

Query: 774  SERLVKEDVYTSVHLEEFESETRDTKLGPEEITREIPNVGEDSLRDLDEMGIIRIGAEVK 833
            SERLVK+DVYTS+H+EEFESE RDTKLGPEE+TR+IPNVGED+LRDLDE GIIR+GAEVK
Sbjct: 779  SERLVKDDVYTSIHIEEFESEARDTKLGPEEMTRDIPNVGEDALRDLDERGIIRVGAEVK 838

Query: 834  EGDILVGKVTPKGEKDLSAEERLLHAIFGDKSREVRDTSLRVPHGGDGVVRDVKIFTRAN 893
            + D+LVGKVTPKG  +L+AEERLLHAIFG+K+REVRDTSLRVPHGG G+V DVKIFTR
Sbjct: 839  DNDLLVGKVTPKGVTELTAEERLLHAIFGEKAREVRDTSLRVPHGGGGIVLDVKIFTREA 898

Query: 894  GDELQSGVNMLVRVYIAQKRKIKVGDKMAGRHGNKGVVSRIVPVEDMPYLPDGTPVDIML 953
            GDEL  GVN LVRVYI QKRKI  GDKMAGRHGNKGV+SRI+P EDMP++PDGTPVDIML
Sbjct: 899  GDELPPGVNQLVRVYIVQKRKIHEGDKMAGRHGNKGVISRILPEEDMPFMPDGTPVDIML 958

Query: 954  NPLGVPSRMNIGQVMELHLGMAARNLGIHIATPVFDGASSEDLWETVQEAGMDSDAKTVL 1013
            NPLGVPSRMNIGQV+ELHLGMAAR LGIH+ATPVFDGA+ ED+W TV+EAGM  DAKT+L
Sbjct: 959  NPLGVPSRMNIGQVLELHLGMAARALGIHVATPVFDGANEEDVWSTVEEAGMARDAKTIL 1018

Query: 1014 YDGRTGEPFDNRVSVGVMYMIKLHHMVDDKLHARSVGPYSLVTQQPLGGKAQFGGQRFGE 1073
            YDGR+GE FDNR+SVGVMYMIKL HMVDDKLHARS GPYSLVTQQPLGGKAQFGGQRFGE
Sbjct: 1019 YDGRSGEAFDNRISVGVMYMIKLAHMVDDKLHARSTGPYSLVTQQPLGGKAQFGGQRFGE 1078

Query: 1074 MEVWALEAYGASNVLQEILTYKSDDVTGRLKAYEAITKGKPIPKPGVPESFRVLVKELQS 1133
            MEVWALEAYGA+  LQEILT KSDDV GR+K YEAI KG+ +P+PGVPESF+VL+KELQS
Sbjct: 1079 MEVWALEAYGAAYTLQEILTIKSDDVVGRVKTYEAIVKGESVPEPGVPESFKVLIKELQS 1138

Query: 1134 LGLDMRVLDEDDNEVELRDLDEGEDDDVMHVDD 1166
            LG+D+++L  D+ E+E+RD+D   DDD  + +D
Sbjct: 1139 LGMDVKMLSADEEEIEMRDMD---DDDFTNQND 1168
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 379> which encodes the amino acid sequence <SEQ ID 380>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3392(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 1129/1190 (94%), Positives = 1168/1190 (97%), Gaps = 3/1190 (0%)

Query: 1    MAGHEVQYGKHRTRRSFSRIKEVLDLPNLIEIQTDSFQDFLDAGLKEVFEDVLPISNFTD 60
            +AGHEV+YGKHRTRRSFSRIKEVLDLPNLIEIQTDSFQDFLD+GLKEVFEDVLPISNFTD
Sbjct: 1    LAGHEVRYGKHRTRRSFSRIKEVLDLPNLIEIQTDSFQDFLDSGLKEVFEDVLPISNFTD 60

Query: 61   TMDLEFVGYELKEPKYTLEEARIHDASYSAPIFVTFRLVNKETGEIKTQEVFFGDFPIMT 120
            TM+LEFVGYE KEPKYTLEEARIHDASYSAPIFVTFRLVNKETGEIKTQEVFFGDFPIMT
Sbjct: 61   TMELEFVGYEFKEPKYTLEEARIHDASYSAPIFVTFRLVNKETGEIKTQEVFFGDFPIMT 120

Query: 121  EMGTFIINGGERIIVSQLVRSPGVYFNDKVDKNGKVGYGSTVIPNRGAWLELETDAKDIA 180
            EMGTFIINGGERIIVSQLVRSPGVYFNDKVDKNGKVGYGSTVIPNRGAWLELETD+KDIA
Sbjct: 121  EMGTFIINGGERIIVSQLVRSPGVYFNDKVDKNGKVGYGSTVIPNRGAWLELETDSKDIA 180

Query: 181  YTRIDRTRKIPFTTLVRALGFSGDDEIVDIFGDSELVRNTIEKDIHKNPSDSRTDEALKE 240
            YTRIDRTRKIPFTTLVRALGFSGDDEIVDIFG+S+LVRNTIEKDIHKNPSDSRTDEALKE
Sbjct: 181  YTRIDRTRKIPFTTLVRALGFSGDDEIVDIFGESDLVRNTIEKDIHKNPSDSRTDEALKE 240

Query: 241  IYERLRPGEPKTADSSRSLLVARFFDPRRYDLAAVGRYKINKKLNLKTRLLNQTIAENLV 300
            IYERLRPGEPKTADSSRSLL+ARFFD RRYDLAAVGRYK+NKKLN+KTRLLNQ IAENLV
Sbjct: 241  IYERLRPGEPKTADSSRSLLIARFFDARRYDLAAVGRYKVNKKLNIKTRLLNQIIAENLV 300
```

```
Query: 301  DGETGEILVEAGTVMTRDVIDSIAEHIDGDLNKFVYTPNDYAVVTEPVILQKFKVVAPTD 360
            D ETGEILVEAGT MTR VI+SI EH+DGDLNKFVYTPNDYAVVTEPV+LQKFKVV+P D
Sbjct: 301  DAETGEILVEAGTEMTRSVIESIEEHLDGDLNKFVYTPNDYAVVTEPVVLQKFKVVSPID 360

Query: 361  PDRVVTIVGNSNPEDKVRALTPADILAEMSYFLNLAEGIGKVDDIDHLGNRRIRAVGELL 420
            PDRVVTIVGN+NP+DKVRALTPADILAEMSYFLNLAEG+GKVDDIDHLGNRRIRAVGELL
Sbjct: 361  PDRVVTIVGNANPDDKVRALTPADILAEMSYFLNLAEGLGKVDDIDHLGNRRIRAVGELL 420

Query: 421  ANQFRIGLARMERNVRERMSVQDNEVLTPQQIINIRPVTAAVKEFFGSSQLSQFMDQHNP 480
            ANQFRIGLARMERNVRERMSVQDN+VLTPQQIINIRPVTAAVKEFFGSSQLSQFMDQHNP
Sbjct: 421  ANQFRIGLARMERNVRERMSVQDNDVLTPQQIINIRPVTAAVKEFFGSSQLSQFMDQHNP 480

Query: 481  LSELSHKRRLSALGPGGLTRDRAGYEVRDVHYTHYGRMCPIETPEGPNIGLINNLSSFGH 540
            LSELSHKRRLSALGPGGLTRDRAGYEVRDVHYTHYGRMCPIETPEGPNIGLINNLSSFGH
Sbjct: 481  LSELSHKRRLSALGPGGLTRDRAGYEVRDVHYTHYGRMCPIETPEGPNIGLINNLSSFGH 540

Query: 541  LNKYGFIQTPYRKVDRSTGAVTNEIVWLTADEEDEFTVAQANSKLNEDGTFAEEIVMGRH 600
            LNKYGFIQTPYRKVDR+TG VTNEIVWLTADEEDE+TVAQANSKLNEDGTFAEEIVMGRH
Sbjct: 541  LNKYGFIQTPYRKVDRATGTVTNEIVWLTADEEDEYTVAQANSKLNEDGTFAEEIVMGRH 600

Query: 601  QGNNQEFPSSIVDFVDVSPKQVVAVATACIPFLENDDSNRALMGANMQRQAVPLIDPKAP 660
            QGNNQEF +S+VDFVDVSPKQVVAVATACIPFLENDDSNRALMGANMQRQAVPLIDPKAP
Sbjct: 601  QGNNQEFSASVVDFVDVSPKQVVAVATACIPFLENDDSNRALMGANMQRQAVPLIDPKAP 660

Query: 661  YVGTGMEYQAAHDSGAAVIAKHDGRVIFSDAEKVEVRREDGSLDVYHVQKFRRSNSGTAY 720
            YVGTGMEYQAAHDSGAAVIA+ +G+V+FSDAEKVE+RR+DGSLDVYH+ KFRRSNSGTAY
Sbjct: 661  YVGTGMEYQAAHDSGAAVIAQQNGKVVFSDAEKVEIRRQDGSLDVYHITKFRRSNSGTAY 720

Query: 721  NQRTLVKVGDLVEKGDFIADGPSMENGEMALGQNPVVAYMTWEGYNFEDAVIMSERLVKE 780
            NQRTLVKVGD+VEKGDFIADGPSMENGEMALGQNPVVAYMTWEGYNFEDAVIMSERLVKE
Sbjct: 721  NQRTLVKVGDIVEKGDFIADGPSMENGEMALGQNPVVAYMTWEGYNFEDAVIMSERLVKE 780

Query: 781  DVYTSVHLEEFESETRDTKLGPEEITREIPNVGEDSLRDLDEMGIIRIGAEVKEGDILVG 840
            DVYTSVHLEEFESETRDTKLGPEEITREIPNVGE++L+DLDEMGIIRIGAEVKEGDILVG
Sbjct: 781  DVYTSVHLEEFESETRDTKLGPEEITREIPNVGEEALKDLDEMGIIRIGAEVKEGDILVG 840

Query: 841  KVTPKGEKDLSAEERLLHAIFGDKSREVRDTSLRVPHGGDGVVRDVKIFTRANGDELQSG 900
            KVTPKGEKDLSAEERLLHAIFGDKSREVRDTSLRVPHGGDG+VRDVKIFTRANGDELQSG
Sbjct: 841  KVTPKGEKDLSAEERLLHAIFGDKSREVRDTSLRVPHGGDGIVRDVKIFTRANGDELQSG 900

Query: 901  VNMLVRVYIAQKRKIKVGDKMAGRHGNKGVVSRIVPVEDMPYLPDGTPVDIMLNPLGVPS 960
            VNMLVRVYIAQKRKIKVGDKMAGRHGNKGVVSRIVPVEDMPYLPDGTPVDIMLNPLGVPS
Sbjct: 901  VNMLVRVYIAQKRKIKVGDKMAGRHGNKGVVSRIVPVEDMPYLPDGTPVDIMLNPLGVPS 960

Query: 961  RMNIGQVMELHLGMAARNLGIHIATPVFDGASSEDLWETVQEAGMDSDAKTVLYDGRTGE 1020
            RMNIGQVMELHLGMAARNLGIHIATPVFDGASSEDLW+TV+EAGMDSDAKTVLYDGRTGE
Sbjct: 961  RMNIGQVMELHLGMAARNLGIHIATPVFDGASSEDLWDTVREAGMDSDAKTVLYDGRTGE 1020

Query: 1021 PFDNRVSVGVMYMIKLHHMVDDKLHARSVGPYSLVTQQPLGGKAQFGGQRFGEMEVWALE 1080
            PFDNRVSVGVMYMIKLHHMVDDKLHARSVGPYSLVTQQPLGGKAQFGGQRFGEMEVWALE
Sbjct: 1021 PFDNRVSVGVMYMIKLHHMVDDKLHARSVGPYSLVTQQPLGGKAQFGGQRFGEMEVWALE 1080

Query: 1081 AYGASNVLQEILTYKSDDVTGRLKAYEAITKGKPIPKPGVPESFRVLVKELQSLGLDMRV 1140
            AYGASNVLQEILTYKSDDVTGRLKAYEAITKGKPIPKPGVPESFRVLVKELQSLGLDMRV
Sbjct: 1081 AYGASNVLQEILTYKSDDVTGRLKAYEAITKGKPIPKPGVPESFRVLVKELQSLGLDMRV 1140

Query: 1141 LDEDDNEVELRDLDEGEDDDVMHVDDLEKARVKQEAEEKQAEQVSEVVQE 1190
            LDEDDNEVELRDLDEGEDDD+MHVDDLEKAR KQ  E    ++VSE   E
Sbjct: 1141 LDEDDNEVELRDLDEGEDDDIMHVDDLEKAREKQAQE---TQEVSETTDE 1187
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 114

A DNA sequence (GBSx0118) was identified in *S.agalactiae* <SEQ ID 381> which encodes the amino acid sequence <SEQ ID 382>. This protein is predicted to be DNA-directed RNA polymerase, beta subunit (rpoC). Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1892(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 383> which encodes the amino acid sequence <SEQ ID 384>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2128(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 1148/1205 (95%), Positives = 1177/1205 (97%)

Query: 11   VVDVNRFKSMQITLASPSKVRSWSYGEVKKPETINYRTLKPEREGLFDEVIFGPTKDWEC 70
            VVDVNRFKSMQITLASPSKVRSWSYGEVKKPETINYRTLKPEREGLFDEVIFGPTKDWEC
Sbjct: 1    VVDVNRFKSMQITLASPSKVRSWSYGEVKKPETINYRTLKPEREGLFDEVIFGPTKDWEC 60

Query: 71   ACGKYKRIRYKGIICDRCGVEVTRAKVRRERMGHIELKAPVSHIWYFKGIPSRMGLTLDM 130
            ACGKYKRIRYKGI+CDRCGVEVTRAKVRRERMGHIELKAPVSHIWYFKGIPSRMGLTLDM
Sbjct: 61   ACGKYKRIRYKGIVCDRCGVEVTRAKVRRERMGHIELKAPVSHIWYFKGIPSRMGLTLDM 120

Query: 131  SPRALEEVIYFAAYVVIDPMDTPLEPKSLLTEREYREKLQEYGYGSFVAKMGAEAIQDLL 190
            SPRALEEVIYFAAYVVIDP DTPLEPKSLLTEREYREKLQEYG+GSFVAKMGAEAIQDLL
Sbjct: 121  SPRALEEVIYFAAYVVIDPKDTPLEPKSLLTEREYREKLQEYGHGSFVAKMGAEAIQDLL 180

Query: 191  KRVDLDAEIAVLKEELKSATGQKRVKAVRRLDVLDAFKKSGNKPEWMVLNILPVIPPDLR 250
            KRVDL AEIA LKEELKSA+GQKR+KAVRRLDVLDAF KSGNKPEWMVLNILPVIPPDLR
Sbjct: 181  KRVDLAAEIAELKEELKSASGQKRIKAVRRLDVLDAFNKSGNKPEWMVLNILPVIPPDLR 240

Query: 251  PMVQLDGGRFAASDLNDLYRRVINRNNRLARLLELNAPGIIVQNEKRMLQEAVDALIDNG 310
            PMVQLDGGRFAASDLNDLYRRVINRNNRLARLLELNAPGIIVQNEKRMLQEAVDALIDNG
Sbjct: 241  PMVQLDGGRFAASDLNDLYRRVINRNNRLARLLELNAPGIIVQNEKRMLQEAVDALIDNG 300

Query: 311  RRGRPITGPGSRPLKSLSHMLKGKQGRFRQNLLGKRVDFSGRSVIAVGPTLKMYQCGVPR 370
            RRGRPITGPGSRPLKSLSHMLKGKQGRFRQNLLGKRVDFSGRSVIAVGPTLKMYQCGVPR
Sbjct: 301  RRGRPITGPGSRPLKSLSHMLKGKQGRFRQNLLGKRVDFSGRSVIAVGPTLKMYQCGVPR 360

Query: 371  EMAIELFKPFVMREIVARDLAGNVKAAKRMVERGDERIWDILEEVIKEHPVLLNRAPTLH 430
            EMAIELFKPFVMREIVA++ AGNVKAAKRMVERGDERIWDILEEVIKEHPVLLNRAPTLH
Sbjct: 361  EMAIELFKPFVMREIVAKEYAGNVKAAKRMVERGDERIWDILEEVIKEHPVLLNRAPTLH 420

Query: 431  RLGIQAFEPVLIDGKALRLHPLVCEAYNADFDGDQMAIHVPLSEEAQAEARLLMLAAEHI 490
            RLGIQAFEPVLIDGKALRLHPLVCEAYNADFDGDQMAIHVPLSEEAQAEARLLMLAAEHI
Sbjct: 421  RLGIQAFEPVLIDGKALRLHPLVCEAYNADFDGDQMAIHVPLSEEAQAEARLLMLAAEHI 480

Query: 491  LNPKDGKPVVTPSQDMVLGNYYLTMEDAGREGEGMIFKDHDEAVMAYQNGYVHLHTRVGI 550
            LNPKDGKPVVTPSQDMVLGNYYLTMEDAGREGEGMIFKD DEAVMAY+NGY HLH+RVGI
Sbjct: 481  LNPKDGKPVVTPSQDMVLGNYYLTMEDAGREGEGMIFKDKDEAVMAYRNGYAHLHSRVGI 540

Query: 551  AVDSMPNKPWTEEQKHKIMVTTVGKILFNDIMPEDLPYLIEPNNANLTEKTPDKYFLEPG 610
```

```
            AVDSMPNKPW + Q+HKIMVTTVGKILFNDIMPEDLPYL EPNNANLTE TPDKYFLEPG
Sbjct: 541  AVDSMPNKPWKDNQRHKIMVTTVGKILFNDIMPEDLPYLQEPNNANLTEGTPDKYFLEPG 600

Query: 611  QDIQAVIDNLEINIPFKKKNLGNIIAETFKRFRTTETSAFLDRLKDLGYYHSTLAGLTVG 670
             QDIQ VID L+IN+PFKKKNLGNIIAETFKRFRTTETSAFLDRLKDLGYYHSTLAGLTVG
Sbjct: 601  QDIQEVIDRLDINVPFKKKNLGNIIAETFKRFRTTETSAFLDRLKDLGYYHSTLAGLTVG 660

Query: 671  IADIPVIDNKAEIIDAAHHRVEDINKAFRRGLMTEEDRYVAVTTTWREAKEALEKRLIET 730
             IADIPVIDNKAEIIDAAHHRVE+INKAFRRGLMT++DRYVAVTTTWREAKEALEKRLIET
Sbjct: 661  IADIPVIDNKAEIIDAAHHRVEEINKAFRRGLMTDDDRYVAVTTTWREAKEALEKRLIET 720

Query: 731  QDPKNPIVMMMDSGARGNISNFSQLAGMRGLMAAPNGRIMELPILSNFREGLSVLEMFFS 790
             QDPKNPIVMMMDSGARGNISNFSQLAGMRGLMAAPNGRIMELPILSNFREGLSVLEMFFS
Sbjct: 721  QDPKNPIVMMMDSGARGNISNFSQLAGMRGLMAAPNGRIMELPILSNFREGLSVLEMFFS 780

Query: 791  THGARKGMTDTALKTADSGYLTRRLVDVAQDVIIREDDCGTDRGLTITAITDGKEVTETL 850
             THGARKGMTDTALKTADSGYLTRRLVDVAQDVIIREDDCGTDRGL I AITDGKEVTETL
Sbjct: 781  THGARKGMTDTALKTADSGYLTRRLVDVAQDVIIREDDCGTDRGLLIRAITDGKEVTETL 840

Query: 851  EERLIGRYTKKSIKHPETGEILVGADTLITEDMAAKVVKAGVEEVTIRSVFTCNTRHGVC 910
             EERL GRYT+KS+KHPETGE+L+GAD LITEDMA K+V AGVEEVTIRSVFTC TRHGVC
Sbjct: 841  EERLQGRYTRKSVKHPETGEVLIGADQLITEDMARKIVDAGVEEVTIRSVFTCATRHGVC 900

Query: 911  RHCYGINLATGDAVEVGEAVGTIAAQSIGEPGTQLTMRTFHTGGVASNTDITQGLPRIQE 970
             RHCYGINLATGDAVEVGEAVGTIAAQSIGEPGTQLTMRTFHTGGVASNTDITQGLPRIQE
Sbjct: 901  RHCYGINLATGDAVEVGEAVGTIAAQSIGEPGTQLTMRTFHTGGVASNTDITQGLPRIQE 960

Query: 971  IFEARNPKGEAVITEVKGEVVAIEEDSSTRTKKVFVKGQTGEGEYVVPFTARMKVEVGDE 1030
             IFEARNPKGEAVITEVKG VV IEED+STRTKKV+V+G+TG GEYV+PFTARMKVEVGDE
Sbjct: 961  IFEARNPKGEAVITEVKGNVVEIEEDASTRTKKVYVQGKTGMGEYVIPFTARMKVEVGDE 1020

Query: 1031 VARGAALTEGSIQPKRLLEVRDTLSVETYLLAEVQKVYRSQGVEIGDKHVEVMVRQMLRK 1090
             V RGAALTEGSIQPKRLLEVRDTLSVETYLLAEVQKVYRSQGVEIGDKHVEVMVRQMLRK
Sbjct: 1021 VNRGAALTEGSIQPKRLLEVRDTLSVETYLLAEVQKVYRSQGVEIGDKHVEVMVRQMLRK 1080

Query: 1091 VRVMDPGDTDLLPGTLMDISDFTDANKDIVISGGIPATSRPVLMGITKASLETNSFLSAA 1150
             VRVMDPGDTDLLPGTLMDISDFTDANKDIVISGGIPATSRPVLMGITKASLETNSFLSAA
Sbjct: 1081 VRVMDPGDTDLLPGTLMDISDFTDANKDIVISGGIPATSRPVLMGITKASLETNSFLSAA 1140

Query: 1151 SFQETTRVLTDAAIRGKKDHLLGLKENVIIGKIIPAGTGMARYRNIEPLAVNEVEIIEGT 1210
             SFQETTRVLTDAAIRGKKDHLLGLKENVIIGKIIPAGTGMARYRNIEP A+NE+E+I+ T
Sbjct: 1141 SFQETTRVLTDAAIRGKKDHLLGLKENVIIGKIIPAGTGMARYRNIEPQAMNEIEVIDHT 1200

Query: 1211 PVDAE 1215
              V AE
Sbjct: 1201 EVSAE 1205
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 115

A DNA sequence (GBSx0120) was identified in *S.agalactiae* <SEQ ID 385> which encodes the amino acid sequence <SEQ ID 386>. This protein is predicted to be a DNA binding protein. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4727(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC45309 GB:U81957 putative DNA binding protein [Streptococcus gordonii]
 Identities = 42/99 (42%), Positives = 75/99 (75%)

Query: 1  MYQVVKMFGDWEPWWFIEGWEEDITEIAEYDTLSEALLYFQEEWDRGQEKWPYFQSKSSL 60
          MY+VV+M+GD+EPWWF++GWE DI +   ++   +AL +++ +W + + ++  ++S+S L
Sbjct: 1  MYRVVEMYGDFEPWWFLDGWENDIIQEQRFEKYYDALKFYKIQWLKLETEFKEYKSRSDL 60

Query: 61 LATFWSIKEKRWCEECDEYLQQYHSLMLLKEWQEIPKEE 99
          +  FW+  ++RWCEECD+Y+QQY S++LL++ + IPK +
Sbjct: 61 MTVFWNENDQRWCEECDDYVQQYRSIILLEDEKVIPKSK 99
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 387> which encodes the amino acid sequence <SEQ ID 388>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.4741(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 61/121 (50%), Positives = 83/121 (68%)

Query: 1   MYQVVKMFGDWEPWWFIEGWEEDITEIAEYDTLSEALLYFQEEWDRGQEKWPYFQSKSSL 60
           MYQV+KM+GDWEPWWFI+GW++DI +   ++    EAL YF +EW R +  +P + S+ +L
Sbjct: 1   MYQVIKMYGDWEPWWFIDGWQDDIIDEQQFSDWQEALDYFNQEWQRMKAIFPSYHSQKNL 60

Query: 61  LATFWSIKEKRWCEECDEYLQQYHSLMLLKEWQEIPKEESIERFEVFNKIAELPSACSLNL 121
           LATFW  ++KRWCE+CDE LQQ+HSL+LLK    +P    I  FE  N   ++   C LNL
Sbjct: 61  LATFWEKEDKRWCEDCDEDLQQFHSLLLLKNKDIVPSNNYIPEFEQRNDSPQVAYLCKLNL 121
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 116

A DNA sequence (GBSx0121) was identified in *S.agalactiae* <SEQ ID 389> which encodes the amino acid sequence <SEQ ID 390>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2433(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC45310 GB:U81957 putative ABC transporter subunit ComYA
          [Streptococcus gordonii]
 Identities = 203/319 (63%), Positives = 255/319 (79%), Gaps = 1/319 (0%)

Query: 1   MVQSLAKQVIHQAVEVNAQDIYIIPKGDCYELYMRIDDERRFIDVFEFNRMASLISHFKF 60
           MVQ +A+ ++ QA E  AQDIY +PK DCYELYMRI DERRFI  ++F+++A++ISHFKF
Sbjct: 1   MVQKIAQAIVRQAKEECAQDIYFVPKDDCYELYMRIGDERRFIQTYDFDQLAAVISHFKF 60

Query: 61  VAGMNVGEKRRSQLGSCDYELSEGRLVSLRLSSVGDYRGQESLVIRILYSGHQDLKYWFD 120
           +AGMNVGEKRRSQLGSCDY   + +  S+RLS+VGDYRG ESLVIR+L+    +LK+WF
Sbjct: 61  LAGMNVGEKRRSQLGSCDYRYDD-KETSIRLSTVGDYRGYESLVIRLLHDEETELKFWFT 119

Query: 121 NIKQMKEVLGIRGLYLFSGPVGSGKTTLMYQLASEVFKNKQIITIEDPVEIKNDKMLQLQ 180
```

```
            +  +++E    RGLYLFSGPVGSGKTTLM+QLA   FK +Q+++IEDPVEIK + MLQLQ
Sbjct: 120 HFPELREKFKDRGLYLFSGPVGSGKTTLMHQLAQLKFKGQQVMSIEDPVEIKQEDMLQLQ 179

Query: 181 LNEDIGMTYDALIKLSLRHRPDILIIGEIRDQATARAVIRASLTGVMVFSTIHAKSIPGV 240
            LNE IG+TY++LIKLSLRHRPD+LIIGEIRD  TARAV+RASLTG  VFSTIHAKSIPGV
Sbjct: 180 LNETIGLTYESLIKLSLRHRPDLLIIGEIRDSETARAVVRASLTGATVFSTIHAKSIPGV 239

Query: 241 YDRLIELGVNYQELENSLKLIAYQRLIGGGSLIDFETGNFKKHSSDKWNRQVDILAEEGH 300
            Y+RL+ELGV+ +EL+  L+ I YQRLIGGG +IDF + N+++H    WN+Q+D L   GH
Sbjct: 240 YERLLELGVSEEELKIVLQGICYQRLIGGGGVIDFASDNYQEHEPTVWNQQIDQLLAAGH 299

Query: 301 ISKKQAQVEKIIPQETTES 319
            I  +QA+ EKI  Q+   S
Sbjct: 300 IHPEQAEAEKIRNQQAKTS 318
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 391> which encodes the amino acid sequence <SEQ ID 392>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1846(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 207/312 (66%), Positives = 257/312 (82%)

Query: 1   MVQSLAKQVIHQAVEVNAQDIYIIPKGDCYELYMRIDDERRFIDVFEFNRMASLISHFKF 60
            MVQ+LAK ++ +A +V+AQDIYI+P+ D Y+L++RI DERR +DV++ +RMA LISHFKF
Sbjct: 1   MVQALAKAILAKAEQVHAQDIYILPRADQYDLFLRIGDERRLVDVYQSDRMAPLISHFKF 60

Query: 61  VAGMNVGEKRRSQLGSCDYELSEGRLVSLRLSSVGDYRGQESLVIRILYSGHQDLKYWFD 120
            VAGM VGEKRR Q+GSCDY+LS+ + +SLRLSSVGDYRGQESLVIR+L+  ++ + YWFD
Sbjct: 61  VAGMIVGEKRRCQVGSCDYKLSKDKQLSLRLSSVGDYRGQESLVIRLLHHQNKSVHYWFD 120

Query: 121 NIKQMKEVLGIRGLYLFSGPVGSGKTTLMYQLASEVFKNKQIITIEDPVEIKNDKMLQLQ 180
             + ++   +G RGLYLF+GPVGSGKTTLMYQL S   +  Q+I+IEDPVEIKN ++LQLQ
Sbjct: 121 GLTKVANQVGGRGLYLFAGPVGSGKTTLMYQLISNYHQEAQVISIEDPVEIKNHQILQLQ 180

Query: 181 LNEDIGMTYDALIKLSLRHRPDILIIGEIRDQATARAVIRASLTGVMVFSTIHAKSIPGV 240
            +N+DIGMTYD LIKLSLRHRPDIL+IGEIRD  TARAVIRASLTG MVFST+HAKSI GV
Sbjct: 181 VNDDIGMTYDNLIKLSLRHRPDILVIGEIRDSQTARAVIRASLTGAMVFSTVHAKSISGV 240

Query: 241 YDRLIELGVNYQELENSLKLIAYQRLIGGGSLIDFETGNFKKHSSDKWNRQVDILAEEGH 300
            Y RL+ELGV   EL N L LIAYQRL+ GG+LID     F+ +SS  WN+Q+D L E GH
Sbjct: 241 YARLLELGVTKAELSNCLALIAYQRLLNGGALIDSTQNEFEYYSSSNWNQQIDQLLEAGH 300

Query: 301 ISKKQAQVEKII 312
            ++ KQA++EKII
Sbjct: 301 LNPKQAKLEKII 312
```

SEQ ID 390 (GBS63) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 5 (lane 5; MW 39kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 13 (lane 2; MW 64kDa).

The GBS63-GST fusion product was purified (Figure 101A; see also Figure 191, lane 3) and used to immunise mice (lane 1 product; 20μg/mouse). The resulting antiserum was used for Western blot (Figure 101B), FACS (Figure 101C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 117

A DNA sequence (GBSx0122) was identified in *S.agalactiae* <SEQ ID 393> which encodes the amino acid sequence <SEQ ID 394>. This protein is predicted to be competence protein (mshG). Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-14.65   Transmembrane  123 - 139 ( 113 - 144)
   INTEGRAL    Likelihood =-13.53   Transmembrane  272 - 288 ( 264 - 295)
   INTEGRAL    Likelihood = -8.55   Transmembrane   79 -  95 (  75 - 102)
   INTEGRAL    Likelihood = -0.00   Transmembrane  146 - 162 ( 146 - 162)

----- Final Results -----
               bacterial membrane --- Certainty=0.6859(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9489> which encodes amino acid sequence <SEQ ID 9490> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC45311 GB:U81957 putative ABC transporter subunit ComYB
          [Streptococcus gordonii]
 Identities = 161/280 (57%), Positives = 219/280 (77%)

Query: 19  MNKALLEGKDLSKMLGELGFSDTVITQVALADLHGNISRSLLKIESYLANLLLVRKKVIE 78
           M + L  G+  S+++  LGFSD V+TQ++LA+LHGN+S +LLKIE YL NL  V+KK+IE
Sbjct: 1   MRQGLANGQAFSEIMASLGFSDAVVTQLSLAELHGNLSLALLKIEEYLDNLAKVKKKLIE 60

Query: 79  VATYPLILLSFLVLIMIGLRNYLMPQLGENNFATRLITNVPNIFLLLLAVVLIFSLIFYI 138
           VATYP++LL FLVLIMIGLRNYL+PQL   NFAT+LI ++P IFLL + ++L  +   Y+
Sbjct: 61  VATYPMMLLGFLVLIMIGLRNYLLPQLSSQNFATQLIGHLPTIFLLTVLMLLGLTGAIYL 120

Query: 139 IQKRLSRIKVACFLTTIPLVGSYVKLYLTAYYAREWGNLLSQGIELDQIVKVMQNQKSKL 198
           + K   RI V  FL  +P VGS+V++YLTAYYAREWGN++ QG+EL QI ++MQ Q+S L
Sbjct: 121 VFKGQKRIPVYSFLARLPFVGSFVRIYLTAYYAREWGNMIGQGLELSQIFQIMQEQRSVL 180

Query: 199 FREIGYDMEEGFLSGKAFHQKVLDYPFFLTELSLMIEYGQVKAKLGTELDIYADEKWEDF 258
           F+EIG D+ +   +G+ F  K+  YPFF  ELSL+IEYG+VK+KLG+EL+IYA + WE+F
Sbjct: 181 FQEIGQDLGQALQNGQEFSDKIASYPFFKKELSLIIEYGEVKSKLGSELEIYALKTWEEF 240

Query: 259 FTKLARATQLIQPVIFIFVALIIVMIYAAMLLPMYQNMEI 298
           F ++ R   LIQP++F+FVAL+IV++YAAMLLP+YQNME+
Sbjct: 241 FGRVNRTMNLIQPLVFVFVALMIVLLYAAMLLPLYQNMEV 280
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 395> which encodes the amino acid sequence <SEQ ID 396>. Analysis of this protein sequence reveals the following:

```
     Possible site: 43
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-12.52   Transmembrane  317 - 333 ( 309 - 339)
   INTEGRAL    Likelihood =-10.14   Transmembrane  123 - 139 ( 119 - 147)
   INTEGRAL    Likelihood = -6.95   Transmembrane  164 - 180 ( 161 - 183)

----- Final Results -----
               bacterial membrane --- Certainty=0.6010(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC45311 GB:U81957 putative ABC transporter subunit ComYB
           [Streptococcus gordonii]
 Identities = 139/278 (50%), Positives = 207/278 (74%)

Query: 63  MEESLLKGQGLADMLSGLGFSDAILTQISLADRHGNIETTLVAIQHYLNQMARIRRKTVE 122
           M + L  GQ  +++++ LGFSDA++TQ+SLA+ HGN+   L+ I+ YL+ +A++++K +E
Sbjct: 1   MRQGLANGQAFSEIMASLGFSDAVVTQLSLAELHGNLSLALLKIEEYLDNLAKVKKKLIE 60

Query: 123 VITYPLILLLFLFVMMLGLRRYLVPQLETQNQITYFLNHFPAFFIGFCSGLILLFGMVWL 182
           V TYP++LL FL ++M+GLR YL+PQL +QN  T  + H P  F+     L+ L G ++L
Sbjct: 61  VATYPMMLLGFLVLIMIGLRNYLLPQLSSQNFATQLIGHLPTIFLLTVLMLLGLTGAIYL 120

Query: 183 RWRSQSRLKLYSRLSRYPFLGKLLKQYLTSYYAREWGTLIGQGLDLMTILDIMAIEKSSL 242
            ++ Q R+ +YS L+R PF+G  ++ YLT+YYAREWG +IGQGL+L  I  IM  ++S L
Sbjct: 121 VFKGQKRIPVYSFLARLPFVGSFVRIYLTAYYAREWGNMIGQGLELSQIFQIMQEQRSVL 180

Query: 243 MKELAEDIRMSLLEGQAFHIKVATYPFFKKELSLMIEYGEIKSKLGAELEIYAQESWEQF 302
            +E+ +D+  +L  GQ F  K+A+YPFFKKELSL+IEYGE+KSKLG+ELEIYA ++WE+F
Sbjct: 181 FQEIGQDLGQALQNGQEFSDKIASYPFFKKELSLIIEYGEVKSKLGSELEIYALKTWEEF 240

Query: 303 FSQLYQVTQLIQPAIFLVVAVTIVMIYAAILLPIYQNM 340
           F ++ +   LIQP +F+ VA+ IV++YAA+LLP+YQNM
Sbjct: 241 FGRVNRTMNLIQPLVFVFVALMIVLLYAAMLLPLYQNM 278
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 148/297 (49%), Positives = 209/297 (69%), Gaps = 2/297 (0%)

Query: 1   MVTFLKRSKLLSDCYTDSMNKALLEGKDLSKMLGELGFSDTVITQVALADLHGNISRSLL 60
           ++ FLKRS+LL   Y   M ++LL+G+ L+ ML  LGFSD ++TQ++LAD HGNI  +L+
Sbjct: 45  VIAFLKRSQLLQLDYVLKMEESLLKGQGLADMLSGLGFSDAILTQISLADRHGNIETTLV 104

Query: 61  KIESYLANLLLVRKKVIEVATYPLILLSFLVLIMIGLRNYLMPQLGENNFATRLITNVPN 120
            I+ YL  +  +R+K +EV TYPLILL FL ++M+GLR YL+PQL   N  T  + + P
Sbjct: 105 AIQHYLNQMARIRRKTVEVITYPLILLLFLFVMMLGLRRYLVPQLETQNQITYFLNHFPA 164

Query: 121 IFL-LLLAVVLIFSLIFYIIQKRLSRIKVACFLTTIPLVGSYVKLYLTAYYAREWGNLLS 179
            F+     ++L+F ++ ++  + SR+K+   L+  P +G  +K YLT+YYAREWG L+
Sbjct: 165 FFIGFCSGLILLFGMV-WLRWRSQSRLKLYSRLSRYPFLGKLLKQYLTSYYAREWGTLIG 223

Query: 180 QGIELDQIVKVMQNQKSKLFREIGYDMEEGFLSGKAFHQKVLDYPFFLTELSLMIEYGQV 239
           QG++L  I+ +M  +KS L +E+  D+    L G+AFH KV  YPFF  ELSLMIEYG++
Sbjct: 224 QGLDLMTILDIMAIEKSSLMKELAEDIRMSLLEGQAFHIKVATYPFFKKELSLMIEYGEI 283

Query: 240 KAKLGTELDIYADEKWEDFFTKLARATQLIQPVIFIFVALIIVMIYAAMLLPMYQNM 296
           K+KLG EL+IYA E WE FF++L + TQLIQP IF+ VA+ IVMIYAA+LLP+YQNM
Sbjct: 284 KSKLGAELEIYAQESWEQFFSQLYQVTQLIQPAIFLVVAVTIVMIYAAILLPIYQNM 340
```

A related GBS gene <SEQ ID 8493> and protein <SEQ ID 8494> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 9
SRCFLG: 0
McG: Length  of  UR:   2
     Peak Value of UR:   1.24
     Net Charge of CR: 0
McG: Discrim Score:     -8.94
GvH: Signal Score (-7.5): -4.08
     Possible site: 31
>>> Seems to have no N-terminal signal sequence
Amino Acid Composition: calculated from 1
ALOM program   count: 4 value: -14.65 threshold:  0.0
   INTEGRAL    Likelihood =-14.65   Transmembrane  105 - 121 (  95 - 126)
   INTEGRAL    Likelihood =-13.53   Transmembrane  254 - 270 ( 246 - 277)
   INTEGRAL    Likelihood = -8.55   Transmembrane   61 -  77 (  57 -  84)
```

```
   PERIPHERAL  Likelihood =  5.09      14
 modified ALOM score:   3.43
icm1 HYPID: 7  CFP: 0.686

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6859(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
57.5/79.7% over 279aa
                                                    Streptococcus gordonii
  GP|2058545| putative ABC transporter subunit ComYB Insert characterized

ORF00008(355 - 1194 of 1500)
GP|2058545|gb|AAC45311.1||U81957(1 - 280 of 282) putative ABC transporter subunit ComYB
{Streptococcus gordonii}
%Match = 33.8
%Identity = 57.5  %Similarity = 79.6
Matches = 161  Mismatches = 57  Conservative Sub.s = 62

144       174       204       234       264       294       324       354
TLRQVILKNTHQTSGIDKWISWLKKDISVRNRHKSKKLSLKKQRKVVQLFNNLFASGFSLTDMVTFLKRSKLLSDCYTDS

384       414       444       474       504       534       564       594
MNKALLEGKDLSKMLGELGFSDTVITQVALADLHGNISRSLLKIESYLANLLLVRKKVIEVATYPLILLSFLVLIMIGLR
| : |  |: :|:::  ||||| |:||::||:||||:| :||||| || ||  |:||:|||||||::|| ||||||||||
MRQGLANGQAFSEIMASLGFSDAVVTQLSLAELHGNLSLALLKIEEYLDNLAKVKKKLIEVATYPMMLLGFLVLIMIGLR
        10        20        30        40        50        60        70        80

624       654       684       714       744       774       804       834
NYLMPQLGENNFATRLITNVPNIFLLLLAVVLIFSLIFYIIQKRLSRIKVACFLTTIPLVGSYVKLYLTAYYAREWGNLL
|||:|||   ||||:|| ::| |||| : ::| ::   |:: |   || |  ||  :|:|||:|::||||||||||||::
NYLLPQLSSQNFATQLIGHLPTIFLLTVLMLLGLTGAIYLVFKGQKRIPVYSFLARLPFVGSFVRIYLTAYYAREWGNMI
        90       100       110       120       130       140       150       160

864       894       924       954       984      1014      1044      1074
SQGIELDQIVKVMQNQKSKLFREIGYDMEEGFLSGKAFHQKVLDYPFFLTELSLMIEYGQVKAKLGTELDIYADEKWEDF
 ||:|| || ::|| |:| ||:||| |: : : :|: |  |:  |||| ||||:||||:||:|||:||:||| : ||:|
GQGLELSQIFQIMQEQRSVLFQEIGQDLGQALQNGQEFSDKIASYPFFKKELSLIIEYGEVKSKLGSELEIYALKTWEEF
       170       180       190       200       210       220       230       240

1104      1134      1164      1194      1224      1254      1284      1314
FTKLARATQLIQPVIFIFVALIIVMIYAAMLLPMYQNMEILS*KIYC*NVRIRRLKHLHF*NVW*HWLQSQELY*FIKD*
| :: |   ||||::|:||||:||::||||||||:|||||:
FGRVNRTMNLIQPLVFVFVALMIVLLYAAMLLPLYQNMEVHL
       250       260       270       280
```

SEQ ID 8494 (GBS49) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 11 (lane 5; MW 15kDa). It was also was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 15 (lane 5; MW 60kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 118

A DNA sequence (GBSx0123) was identified in *S.agalactiae* <SEQ ID 397> which encodes the amino acid sequence <SEQ ID 398>. This protein is predicted to be ComYD or ComGD. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA75315 GB:Y15043 homology to ComYD from Streptcoccus gordonii,
          and ComGD from Bacillus subtilis [Lactococcus lactis subsp. cremoris]
 Identities = 56/138 (40%), Positives = 92/138 (66%), Gaps = 2/138 (1%)

Query: 12  KVKAFTLLECLVALVTITGALLVYQGLTKLLAQQIVVMSSSSQSEWVLLTQQLNAEFEGA 71
           K++AFTLLECLVAL+ I+G++LV  GLT+++ +Q+ +  + S+ +W +  +Q+ +E  GA
Sbjct: 13  KIRAFTLLECLVALLAISGSVLVISGLTRMIEEQMKISQNDSRKDWQIFCEQMRSELSGA 72

Query: 72  HLEYLRQNKLYLRKQDKIVTFGKSNKDDFRKTGYDGRGYQPMVYGLDNCQMSQTKSMVKL 131
            L+ + QN LY+ K DK + FG    DDFRK+   G+GYQPM+Y L   ++   ++++K+
Sbjct: 73  KLDNVNQNFLYVTK-DKKLRFGLVG-DDFRKSDDKGQGYQPMLYDLKGAKIQAEENLIKI 130

Query: 132 VFYFKDGLKRTFYYDFKE 149
              F +G +R F Y F +
Sbjct: 131 TIDFDNGGERVFIYRFTD 148
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 399> which encodes the amino acid sequence <SEQ ID 400>. Analysis of this protein sequence reveals the following:

```
     Possible site: 28

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA75315 GB:Y15043 homology to ComYD from Streptcoccus gordonii,
          and ComGD from Bacillus subtilis [Lactococcus lactis subsp. cremoris]
 Identities = 65/137 (47%), Positives = 84/137 (60%), Gaps = 2/137 (1%)

Query: 8   IKAFTLLEALIALLVISGSLLVYQGLTRTLLKHSHYLARHDQDNWLLFSHQLREELSGAR 67
           I+AFTLLE L+ALL ISGS+LV  GLTR + +          + +W +F  Q+R ELSGA+
Sbjct: 14  IRAFTLLECLVALLAISGSVLVISGLTRMIEEQMKISQNDSRKDWQIFCEQMRSELSGAK 73

Query: 68  FYKVADNKLYVEKGKKVLAFGQFKSHDFRKSASNGKGYQPMLFGISRSHIHIEQSQICIT 127
              V  N LYV K KK L FG     DFRKS   G+GYQPML+ +  + I  E++ I IT
Sbjct: 74  LDNVNQNFLYVTKDKK-LRFG-LVGDDFRKSDDKGQGYQPMLYDLKGAKIQAEENLIKIT 131

Query: 128 LKWKSGLERTFYYAFQD 144
           + + +G ER F Y F D
Sbjct: 132 IDFDNGGERVFIYRFTD 148
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 58/137 (42%), Positives = 88/137 (63%)

Query: 13  VKAFTLLECLVALVTITGALLVYQGLTKLLAQQIVVMSSSSQSEWVLLTQQLNAEFEGAH 72
           +KAFTLLE L+AL+ I+G+LLVYQGLT+ L +    ++   Q  W+L + QL  E  GA
Sbjct: 8   IKAFTLLEALIALLVISGSLLVYQGLTRTLLKHSHYLARHDQDNWLLFSHQLREELSGAR 67

Query: 73  LEYLRQNKLYLRKQDKIVTFGKSNKDDFRKTGYDGRGYQPMVYGLDNCQMSQTKSMVKLV 132
              +  NKLY+ K  K++ FG+    DFRK+  +G+GYQPM++G+    +S + + + LV
Sbjct: 68  FYKVADNKLYVEKGKKVLAFGQFKSHDFRKSASNGKGYQPMLFGISRSHIHIEQSQICIT 127
```

```
Query: 133 FYFKDGLKRTFYYDFKE 149
            +K GL+RTFYY F++
Sbjct: 128 LKWKSGLERTFYYAFQD 144
```

A related GBS gene <SEQ ID 8495> and protein <SEQ ID 8496> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:      4.86
GvH: Signal Score (-7.5): -0.22
     Possible site: 55
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:  12.47 threshold:  0.0
   PERIPHERAL  Likelihood = 12.47     127
 modified ALOM score:  -2.99

*** Reasoning Step: 3

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
GP|3287181| homology to ComYD from Streptcoccus gordonii, and ComGD from Bacillus subtilis
{Lactococcus lactis subsp. cremoris} Inse
rt characterized

ORF00009(334 - 747 of 1053)
GP|3287181|emb|CAA75315.1||Y15043(13 - 148 of 150) homology to ComYD from Streptcoccus
gordonii, and ComGD from Bacillus subtilis {L
actococcus lactis subsp. cremoris}
%Match = 15.9
%Identity = 40.6  %Similarity = 68.1
Matches = 56  Mismatches = 42  Conservative Sub.s = 38

177       207       237       267       297       327       357       387
IC**EVGGFFYKIS*SDPVNPTRYFYFCSSYHCYDLCSNAVTNVSKYGDIIMKNLLLKCKDKKVKAFTLLECLVALVTIT
                                                  :     :     |::|||||||||||: |:
                                                  MTMERKFCDLKLKIRAFTLLECLVALLAIS
                                                          10        20        30

417       447       477       507       537       567       597       627
GALLVYQGLTKLLAQQIVVMSSSSQSEWVLLTQQLNAEFEGAHLEYLRQNKLYLRKQDKIVTFGKSNKDDFRKTGYDGRG
|::||  |||::: :|: :  : |: :| :: :|: :|: || |: : || ||: | || : ||    |||||:   |:|
GSVLVISGLTRMIEEQMKISQNDSRKDWQIFCEQMRSELSGAKLDNVNQNFLYVTK-DKKLRFGLVG-DDFRKSDDKGQG
        40        50        60        70        80        90        100

657       687       717       747       777       807       837       867
YQPMVYGLDNCQMSQTKSMVKLVFYFKDGLKRTFYYDFKEET*SWHPFASYCIGCCIYTRLTVLSSKNIGNRKTVS*PN*
||||:| |   ::   ::::|:   | :| :| | | | :
YQPMLYDLKGAKIQAEENLIKITIDFDNGGERVFIYRFTDTK
         120       130       140       150
```

SEQ ID 398 (GBS6) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 1 (lane 2; MW 40kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 2 (lane 2; MW 15kDa). The GBS6-GST fusion product was purified (Figure 189, lane 2) and used to immunise mice. The resulting antiserum was used for FACS (Figure 260), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 119

A DNA sequence (GBSx0124) was identified in *S.agalactiae* <SEQ ID 401> which encodes the amino acid sequence <SEQ ID 402>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3831(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC00317 GB:AF008220 YtxK [Bacillus subtilis]
 Identities = 106/329 (32%), Positives = 176/329 (53%), Gaps = 17/329 (5%)

Query: 1   MNFEKIETAYELILENIQTIENQLKTHIYDALIEQNSYYLGSSCDLDMVVVNNQKLRQLD 60
           M  + +   YEL+ E    I+N+L+    +AL E    Y     D  + +  +QK +QL
Sbjct: 1   MQKDHVGAVYELLNEAAIMIKNELQISYIEALAEAGEMYFLEKTD-QLKLPADQKTKQLQ 59

Query: 61  LSQE---------EW-RRTFQFIFIKSAQTEQLQANHQFTPDSIGFILLFLLEE-LTSQE 109
              E          EW R+ FQ   +K  + +    N Q TPD+IG  + +L+ + +  ++
Sbjct: 60  ALLEKAEFGTYEHEWVRKAFQLAVLKGMK-DISHPNRQMTPDTIGLFISYLVNKFMADKK 118

Query: 110 TVDVLEIGSGTGNLAQTLLNN-SSKELNYMGIEVDDLLIDLSASIAEIIGSSAQFIQEDA 168
            + +L+   GTGNL  T+LN  S K  N  GIE+DD+L+ ++ + A ++    +   +D+
Sbjct: 119 ELTILDPALGTGNLLFTVLNQLSEKTANSFGIEIDDVLLKIAYAQANLLKKELELFHQDS 178

Query: 169 VRPQILKESDVIISDLPVGYYPNDGIAKRYAVSSSKEHTYAHHLLMEQSLKYLKKDGIAI 228
           + P  +   D +I DLPVGYYPND  A+ + + + + H++AHHL +EQS+K+ K  G
Sbjct: 179 LEPLFIDPVDTVICDLPVGYYPNDEGAEAFELKADEGHSFAHHLFIEQSVKHTKPGGYLF 238

Query: 229 FLAPENLLTSPQSDLLKEWLKGYADVIAVLTLPETIFGSRQNAKSIFVLKKQAEQKP--- 285
           F+ P +L  S QS  LK++ K    + A+L LP++IF    +AKSI VL+KQ E
Sbjct: 239 FMIPNHLFESSQSGKLKQFFKDKVHINALLQLPKSIFKDEAHAKSILVLQKQGENTKAPG 298

Query: 286 ETFVYPLTDLQNRENMANFIENFQKWSRE 314
           +  +  L    N++ M + +  F +W ++
Sbjct: 299 QILLANLPSFSNQKAMLDMMAQFDEWFKK 327
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 403> which encodes the amino acid sequence <SEQ ID 404>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 223/315 (70%), Positives = 270/315 (84%)

Query: 1   MNFEKIETAYELILENIQTIENQLKTHIYDALIEQNSYYLGSSCDLDMVVVNNQKLRQLD 60
           M FEKIE AY+L+LEN Q IEN LKTHIYDA++EQNS+YLG+      V  N+ KL+ L
Sbjct: 16  MTFEKIEEAYQLLLENCQLIENDLKTHIYDAIVEQNSFYLGAEGASPQVAQNSDKLKALC 75

Query: 61  LSQEEWRRTFQFIFIKSAQTEQLQANHQFTPDSIGFILLFLLEELTSQETVDVLEIGSGT 120
```

```
           L++EEWR+ +QF+FIK+AQTEQLQANHQFTPD+IGFILL+LLE+L+ +++++VLEIGSGT
Sbjct: 76  LTKEEWRKAYQFLFIKAAQTEQLQANHQFTPDAIGFILLYLLEQLSDKDSLEVLEIGSGT 135

Query: 121 GNLAQTLLNNSSKELNYMGIEVDDLLIDLSASIAEIIGSSAQFIQEDAVRPQILKESDVI 180
           GNLAQTLLNN+SK L+Y+GIE+DDLLIDLSASIAEI+ SSA FIQEDAVRPQ+LKESD++
Sbjct: 136 GNLAQTLLNNTSKSLDYVGIELDDLLIDLSASIAEIMDSSAHFIQEDAVRPQLLKESDIV 195

Query: 181 ISDLPVGYYPNDGIAKRYAVSSSKEHTYAHHLLMEQSLKYLKKDGIAIFLAPENLLTSPQ 240
           ISDLPVGYYPND IAKRY V+SS +HTYAHHLLMEQSLKYLKKDG AIFLAP NLLTSPQ
Sbjct: 196 ISDLPVGYYPNDDIAKRYKVASSDKHTYAHHLLMEQSLKYLKKDGFAIFLAPVNLLTSPQ 255

Query: 241 SDLLKEWLKGYADVIAVLTLPETIFGSRQNAKSIFVLKKQAEQKPETFVYPLTDLQNREN 300
           S LLK+WLK YA V+ ++TLP++IFG   NAKSI VL+KQ +   ETFVYP+ DL+  EN
Sbjct: 256 SQLLKQWLKDYAQVVTLITLPDSIFGHPSNAKSIIVLQKQTDHPMETFVYPIRDLKLAEN 315

Query: 301 MANFIENFQKWSREN 315
           + +F+ENF+KW   N
Sbjct: 316 IHDFMENFKKWKLSN 330
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 120

A DNA sequence (GBSx0125) was identified in *S.agalactiae* <SEQ ID 405> which encodes the amino acid sequence <SEQ ID 406>. This protein is predicted to be acetate kinase (ackA-1). Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2384(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC36857 GB:L17320 acetate kinase [Bacillus subtilis]
 Identities = 223/395 (56%), Positives = 293/395 (73%), Gaps = 3/395 (0%)

Query: 1   MSKTIAINAGSSSLKWQLYEMPEEKVVAKGIIERIGLKDSISTVKFDDKKDEQILDIVDH 60
           MSK IAINAGSSSLK+QL+EMP E V+ KG++ERIG+ DS+ T+  + +K+ ++ DI DH
Sbjct: 1   MSKIIAINAGSSSLKFQLFEMPSETVLTKGLVERIGIADSVFTISVNGEKNTEVTDIPDH 60

Query: 61  TQAVKILLEDLTKHGIIKDFNEITGVGHRVVAGGEYFKESALVDDKVVEQVEELSALAPL 120
             AVK+LL  LT+ GIIKD NEI G+GHRVV GGE F +S L+ D+ ++++E++S LAPL
Sbjct: 61  AVAVKMLLNKLTEFGIIKDLNEIDGIGHRVVHGGEKFSDSVLLTDETIKEIEDISELAPL 120

Query: 121 HNPAAAAGIRAFREILPDITSVCVFDTAFHTTMQPHTYLYPIPQKYYTDYKVRKYGAHGT 180
           HNPA   GI+AF+E+LP++ +V VFDTAFH TM   +YLY +P +YY  + +RKYG HGT
Sbjct: 121 HNPANIVGIKAFKEVLPNVPAVAVFDTAFHQTMPEQSYLYSLPYEYYEKFGIRKYGFHGT 180

Query: 181 SHQYVAQEAAKQLGRPLEELKLITAHVGNGVSITANYHGQSIDTSMGFTPLAGPMMGTRS 240
           SH+YV + AA+ LGRPL++L+LI+ H+GNG SI A   G+SIDTSMGFTPLAG  MGTRS
Sbjct: 181 SHKYVTERAAELLGRPLKDLRLISCHLGNGASIAAVEGGKSIDTSMGFTPLAGVAMGTRS 240

Query: 241 GDIDPAIIPYLVANDPELEDAAAVVNMLNKQSGLLGVSGTSSDMRDIEAGLQSKDPNAVL 300
           G+IDPA+IPY++    +  D   V+N LNK+SGLLG+SG SSD+RDI   +  +  A
Sbjct: 241 GNIDPALIPYIMEKTGQTAD--EVLNTLNKKSGLLGISGFSSDLRDIVEATKEGNERAET 298

Query: 301 AYNVFIDRIKKFIGQYLAVLNGADAIIFTAGMGENAPLMRQDVIAGLSWFGIELDPE-KN 359
           A  VF  RI K+IG Y A ++G DAIIFTAG+GEN+  +R+ V+ GL + G+  DP   N
Sbjct: 299 ALEVFASRIHKYIGSYAARMSGVDAIIFTAGIGENSVEVRERVLRGLEFMGVYWDPALNN 358
```

```
Query: 360 VFGYFGDITKPDSKVKVLVIPTDEELMIARDVERL 394
           V G    I+ P S VKV++IPTDEE+MIARDV RL
Sbjct: 359 VRGEEAFISYPHSPVKVMIIPTDEEVMIARDVVRL 393
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 407> which encodes the amino acid sequence <SEQ ID 408>. Analysis of this protein sequence reveals the following:

```
     Possible site: 28

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.22   Transmembrane   63 -  79 (  63 -  79)

----- Final Results -----
              bacterial membrane --- Certainty=0.1086(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC36857 GB:L17320 acetate kinase [Bacillus subtilis]
 Identities = 218/395 (55%), Positives = 293/395 (73%), Gaps = 3/395 (0%)

Query: 1   MSKTIAINAGSSSLKWQLYQMPEEAVLAQGIIERIGLKDSISTVKYDGKKEEQILDIHDH 60
           MSK IAINAGSSSLK+QL++MP E VL +G++ERIG+ DS+ T+  +G+K  ++ DI DH
Sbjct: 1   MSKIIAINAGSSSLKFQLFEMPSETVLTKGLVERIGIADSVFTISVNGEKNTEVTDIPDH 60

Query: 61  TEAVKILLNDLIHFGIIAAYDEITGVGHRVVAGGELFKESVVVNDKVLEQIEELSVLAPL 120
             AVK+LLN L  FGII   +EI G+GHRVV GGE F +SV++ D+ +++IE++S LAPL
Sbjct: 61  AVAVKMLLNKLTEFGIIKDLNEIDGIGHRVVHGGEKFSDSVLLTDETIKEIEDISELAPL 120

Query: 121 HNPGAAAGIRAFRDILPDITSVCVFDTSFHTSMAKHTYLYPIPQKYYTDYKVRKYGAHGT 180
           HNP    GI+AF+++LP++ +V VFDT+FH +M + +YLY +P +YY  + +RKYG HGT
Sbjct: 121 HNPANIVGIKAFKEVLPNVPAVAVFDTAFHQTMPEQSYLYSLPYEYYEKFGIRKYGFHGT 180

Query: 181 SHKYVAQEAAKMLGRPLEELKLITAHIGNGVSITANYHGKSVDTSMGFTPLAGPMMGTRS 240
           SHKYV + AA++LGRPL++L+LI+ H+GNG SI A   GKS+DTSMGFTPLAG  MGTRS
Sbjct: 181 SHKYVTERAAELLGRPLKDLRLISCHLGNGASIAAVEGGKSIDTSMGFTPLAGVAMGTRS 240

Query: 241 GDIDPAIIPYLIEQDPELKDAADVVNMLNKKSGLSGVSGISSDMRDIEAGLQEDNPDAVL 300
           G+IDPA+IPY++E+  +  D  +V+N LNKKSGL G+SG SSD+RDI    +E N  A
Sbjct: 241 GNIDPALIPYIMEKTGQTAD--EVLNTLNKKSGLLGISGFSSDLRDIVEATKEGNERAET 298

Query: 301 AYNIFIDRIKKCIGQYFAVLNGADALVFTAGMGENAPLMRQDVIGGLTWFGMDIDPE-KN 359
           A  +F  RI K IG Y A ++G DA++FTAG+GEN+  +R+ V+ GL + G+  DP   N
Sbjct: 299 ALEVFASRIHKYIGSYAARMSGVDAIIFTAGIGENSVEVRERVLRGLEFMGVYWDPALNN 358

Query: 360 VFGYRGDISTPESKVKVLVISTDEELCIARDVERL 394
           V G    IS P S VKV++I TDEE+ IARDV RL
Sbjct: 359 VRGEEAFISYPHSPVKVMIIPTDEEVMIARDVVRL 393
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 332/395 (84%), Positives = 365/395 (92%)

Query: 1   MSKTIAINAGSSSLKWQLYEMPEEKVVAKGIIERIGLKDSISTVKFDDKKDEQILDIVDH 60
           MSKTIAINAGSSSLKWQLY+MPEE V+A+GIIERIGLKDSISTVK+D KK+EQILDI DH
Sbjct: 1   MSKTIAINAGSSSLKWQLYQMPEEAVLAQGIIERIGLKDSISTVKYDGKKEEQILDIHDH 60

Query: 61  TQAVKILLEDLTKHGIIKDFNEITGVGHRVVAGGEYFKESALVDDKVVEQVEELSALAPL 120
           T+AVKILL DL   GII  ++EITGVGHRVVAGGE FKES +V+DKV+EQ+EELS LAPL
Sbjct: 61  TEAVKILLNDLIHFGIIAAYDEITGVGHRVVAGGELFKESVVVNDKVLEQIEELSVLAPL 120

Query: 121 HNPAAAAGIRAFREILPDITSVCVFDTAFHTTMQPHTYLYPIPQKYYTDYKVRKYGAHGT 180
           HNP AAAGIRAFR+ILPDITSVCVFDT+FHT+M  HTYLYPIPQKYYTDYKVRKYGAHGT
Sbjct: 121 HNPGAAAGIRAFRDILPDITSVCVFDTSFHTSMAKHTYLYPIPQKYYTDYKVRKYGAHGT 180

Query: 181 SHQYVAQEAAKQLGRPLEELKLITAHVGNGVSITANYHGQSIDTSMGFTPLAGPMMGTRS 240
```

```
           SH+YVAQEAAK LGRPLEELKLITAH+GNGVSITANYHG+S+DTSMGFTPLAGPMMGTRS
Sbjct: 181 SHKYVAQEAAKMLGRPLEELKLITAHIGNGVSITANYHGKSVDTSMGFTPLAGPMMGTRS 240

Query: 241 GDIDPAIIPYLVANDPELEDAAAVVNMLNKQSGLLGVSGTSSDMRDIEAGLQSKDPNAVL 300
           GDIDPAIIPYL+  DPEL+DAA VVNMLNK+SGL GVSG SSDMRDIEAGLQ  +P+AVL
Sbjct: 241 GDIDPAIIPYLIEQDPELKDAADVVNMLNKKSGLSGVSGISSDMRDIEAGLQEDNPDAVL 300

Query: 301 AYNVFIDRIKKFIGQYLAVLNGADAIIFTAGMGENAPLMRQDVIAGLSWFGIELDPEKNV 360
           AYN+FIDRIKK IGQY AVLNGADA++FTAGMGENAPLMRQDVI GL+WFG+++DPEKNV
Sbjct: 301 AYNIFIDRIKKCIGQYFAVLNGADALVFTAGMGENAPLMRQDVIGGLTWFGMDIDPEKNV 360

Query: 361 FGYFGDITKPDSKVKVLVIPTDEELMIARDVERLK 395
           FGY GDI+ P+SKVKVLVI TDEEL IARDVERLK
Sbjct: 361 FGYRGDISTPESKVKVLVISTDEELCIARDVERLK 395
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 121

A DNA sequence (GBSx0126) was identified in *S.agalactiae* <SEQ ID 409> which encodes the amino acid sequence <SEQ ID 410>. This protein is predicted to be repressor protein. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB49550 GB:AJ248284 repressor protein, putative [Pyrococcus
          abyssi]
 Identities = 39/64 (60%), Positives = 49/64 (75%)

Query: 1  MKNSLQKLRKSRKLSQAELAVALGVTRQTIISLEKEKYTASLELAFKIARYFDKQIEEVF 60
           MKN L++ R+   L+Q ELA  LGVTRQTII++EK KY  SL LAFKIAR+F  +IE++F
Sbjct: 1  MKNRLREFREKYGLTQEELARILGVTRQTIIAIEKGKYDPSLRLAFKIARFFGVRIEDIF 60

Query: 61 IYTE 64
           IY E
Sbjct: 61 IYEE 64
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 411> which encodes the amino acid sequence <SEQ ID 412>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4344(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 29/66 (43%), Positives = 44/66 (65%)

Query: 1  MKNSLQKLRKSRKLSQAELAVALGVTRQTIISLEKEKYTASLELAFKIARYFDKQIEEVF 60
           +KN L++LR    ++Q E+A   GV+RQTI  +E+ +YT S+ +A KIA+ F + +EEVF
Sbjct: 10 LKNRLKELRARDGINQTEMAKLAGVSRQTISLIERNEYTPSVIIAMKIAKVFQEPVEEVF 69
```

```
Query: 61 IYTESE 66
          E E
Sbjct: 70 RLVEVE 75
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 122

A DNA sequence (GBSx0127) was identified in *S.agalactiae* <SEQ ID 413> which encodes the amino acid sequence <SEQ ID 414>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.97   Transmembrane   45 -  61 (  41 -  66)
   INTEGRAL    Likelihood = -8.65   Transmembrane   14 -  30 (  11 -  37)
   INTEGRAL    Likelihood = -7.80   Transmembrane  123 - 139 ( 118 - 145)
   INTEGRAL    Likelihood = -3.24   Transmembrane  177 - 193 ( 177 - 194)
   INTEGRAL    Likelihood = -0.85   Transmembrane   81 -  97 (  81 -  97)

----- Final Results -----
               bacterial membrane --- Certainty=0.4588(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9491> which encodes amino acid sequence <SEQ ID 9492> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA11325 GB:D78257 ORF8 [Enterococcus faecalis]
 Identities = 48/120 (40%), Positives = 69/120 (57%), Gaps = 5/120 (4%)

Query: 104 MQGVKDTANQTVIMELTKQLPLALMLIFAIIGAPIMEEIIFRYIIPKELFAKHQKWGFVI 163
           MQG   TAN + +++L   +   L+++   I APIMEEI+FR  I   L   +     +I
Sbjct: 1   MQGHTTTANDSTLIKLFSGVSPVLVVLLLGIAAPIMEEIVFRGGIIGYLVENNALLAILI 60

Query: 164 GTLAFALIHSPSDIGSFIIYAGMGAILSFVYYKTEHLEYSIMIHFINN-----ALAYSVL 218
            +  F +IH P++  SF +Y  MG ILS  YYKT+ L  SI IHF+NN     A+AY ++
Sbjct: 61  SSFLFGIIHGPTNFISFGMYFFMGIILSVSYYKTKDLRVSISIHFLNNLFPAIAIAYGLI 120
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 415> which encodes the amino acid sequence <SEQ ID 416>. Analysis of this protein sequence reveals the following:

```
     Possible site: 24
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.41   Transmembrane   12 -  28 (   1 -  30)
   INTEGRAL    Likelihood = -9.98   Transmembrane   41 -  57 (  33 -  64)
   INTEGRAL    Likelihood = -8.33   Transmembrane  128 - 144 ( 121 - 151)
   INTEGRAL    Likelihood = -7.96   Transmembrane   83 -  99 (  76 - 103)
   INTEGRAL    Likelihood = -3.77   Transmembrane  208 - 224 ( 207 - 230)
   INTEGRAL    Likelihood = -2.13   Transmembrane  182 - 198 ( 182 - 199)

----- Final Results -----
               bacterial membrane --- Certainty=0.5564(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAA11325 GB:D78257 ORF8 [Enterococcus faecalis]
 Identities = 47/120 (39%), Positives = 70/120 (58%), Gaps = 8/120 (6%)
```

```
Query: 105 GQQVSANDAAIHTLARLIKGGFPLYTALFVLVIAFIAPIMEELVFRGFPMIDLFKGKSLK 164
           G   +AND+   TL +L  G  P+   L VL++   APIMEE+VFRG  +  L +  +L
Sbjct: 3   GHTTTANDS---TLIKLFSGVSPV---LVVLLLGIAAPIMEEIVFRGGIIGYLVENNAL- 55

Query: 165 VAGLVTSLVFALPHA-TNSVEFIMYSCMGIFLFVAYQRRGNLKDAILLHIFNNLIEVILL 223
           +A L++S +F + H  TN + F MY  MGI L V+Y +  +L+ +I +H  NNL   I +
Sbjct: 56  LAILISSFLFGIIHGPTNFISFGMYFFMGIILSVSYYKTKDLRVSISIHFLNNLFPAIAI 115
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 72/229 (31%), Positives = 114/229 (49%), Gaps = 24/229 (10%)

Query: 11  KGKILALLIAFLVINQLV-PILAVWLLKNHYQTPFTSILLIGL-------ELLIIALFLY 62
           KG I  L IA L+I  +V  +L + LL+  +  P      IG+        +LI+   LY
Sbjct: 2   KGFINYLKIAVLIILAMVFNVLPMILLQKQHDIPMVLNWGIGIFYLVIVGSVLIVLWGLY 61

Query: 63  YAKVKQIIRWKALLTRKALVT---ILLGWLSLRVPQIIGYLIMTM-QGVKDTANQTVIME 118
            AK    I+ + +     LV    + L WL +RV  I+G L+  +  G + +AN   I
Sbjct: 62  QAKQDTFIKQQKM----RLVDWGYLALFWLIIRVIAIVGTLVNQLWSGQQVSANDAAIHT 117

Query: 119 LTKQL----PLALMLIFAIIG--APIMEEIIFRYIIPKELF-AKHQKWGFVIGTLAFALI 171
           L + +    PL   L   +I   APIMEE++FR     +LF  K  K   ++ +L FAL
Sbjct: 118 LARLIKGGFPLYTALFVLVIAFIAPIMEELVFRGFPMIDLFKGKSLKVAGLVTSLVFALP 177

Query: 172 HSPSDIGSFIIYAGMGAILSFVYYKTEHLEYSIMIHFINNALAYSVLIS 220
           H+ + +  FI+Y+ MG  L   Y +  +L+ +I++H  NN +   +L+S
Sbjct: 178 HATNSV-EFIMYSCMGIFLFVAYQRRGNLKDAILLHIFNNLIEVILLMS 225
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 123

A DNA sequence (GBSx0128) was identified in *S.agalactiae* <SEQ ID 417> which encodes the amino acid sequence <SEQ ID 418>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0826(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC06504 GB:AE000676 pyrroline carboxylate reductase [Aquifex
            aeolicus]
 Identities = 97/259 (37%), Positives = 159/259 (60%), Gaps = 4/259 (1%)

Query: 1   MKIGIIGVGKM--ASAIIQGLKQTQHDIIISGSCLERSKEIAERLDVTYAESHQSLINQA 58
           M++GI+G G M  A A+    K  + +II++    E+ + +A  + + +A   + L + +
Sbjct: 8   MRVGIVGFGNMGQAFALCFSKKLGKENIIVTDKVQEK-RNLATEMGIAFASDVKFLADNS 66

Query: 59  DIIMLGIKPQLFEKVLLPLDITKPII-SMAAGISLARLSQLTRSDLPLIRIMPNINAQIL 117
           D++++ +KP+  ++VL  L   K II S+ AG+S+ ++ ++   D  ++R+MPN+N  +
Sbjct: 67  DVVLVAVKPKDSQEVLQKLKDYKGIILSIMAGVSIEKMEKILGKDKKIVRVMPNVNVAVG 126

Query: 118 QSCTAICYNNHVSDELRQLAKEITDSFGSSFDIAETNFDTFTALAGSSPAYIYLFIEALA 177
               AI  N ++S+E R   +E+  S G+ + I E  FD FTALAGS PA+++ FI+ALA
Sbjct: 127 SGVMAITDNGNLSEEERSKVEELLLSCGTLYRIEERLFDAFTALAGSGPAFVFSFIDALA 186

Query: 178 KAGVKYGFPKEQALSIVGQTVLASSQNLLQGQNSTSDLIDNICSPGGTTIAGLLDLEKNG 237
            AGV  GF  EQAL I   TV+ S++ L + Q + ++LI  + SPGGTTI G+  LE+ G
Sbjct: 187 LAGVHQGFSYEQALRIALDTVMGSAKLLKEFQVNPNELIAKVTSPGGTTIEGIKYLEEKG 246
```

```
Query: 238 LTHSVISAIDATIEKAKKL 256
              +V+  I+ T +KAKKL
Sbjct: 247 FKGTVMECINRTSQKAKKL 265
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 419> which encodes the amino acid sequence <SEQ ID 420>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1043(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 180/256 (70%), Positives = 208/256 (80%)

Query: 1   MKIGIIGVGKMASAIIQGLKQTQHDIIISGSCLERSKEIAERLDVTYAESHQSLINQADI 60
           MKIGIIGVGKMASAII+GLKQT H++IISGS LERSKEIAE+L + YA SHQ LI+Q D+
Sbjct: 1   MKIGIIGVGKMASAIIKGLKQTPHELIISGSSLERSKEIAEQLALPYAMSHQDLIDQVDL 60

Query: 61  IMLGIKPQLFEKVLLPLDITKPIISMAAGISLARLSQLTRSDLPLIRIMPNINAQILQSC 120
           ++LGIKPQLFE VL PL   +PIISMAAGISL RL+     DLPL+RIMPN+NAQILQS
Sbjct: 61  VILGIKPQLFETVLKPLHFKQPIISMAAGISLQRLATFVGQDLPLLRIMPNMNAQILQSS 120

Query: 121 TAICYNNHVSDELRQLAKEITDSFGSSFDIAETNFDTFTALAGSSPAYIYLFIEALAKAG 180
           TA+  N  VS EL+   +++TDSFGS+FDI+E +FDTFTALAGSSPAYIYLFIEALAKAG
Sbjct: 121 TALTGNALVSQELQARVRDLTDSFGSTFDISEKDFDTFTALAGSSPAYIYLFIEALAKAG 180

Query: 181 VKYGFPKEQALSIVGQTVLASSQNLLQGQNSTSDLIDNICSPGGTTIAGLLDLEKNGLTH 240
           VK G PK +AL IV QTVLAS+ NL     S  D ID ICSPGGTTIAGL++LE+ GLT
Sbjct: 181 VKNGIPKAKALEIVTQTVLASASNLKTSSQSPHDFIDAICSPGGTTIAGLMELERLGLTA 240

Query: 241 SVISAIDATIEKAKKL 256
           +V SAID TI+KAK L
Sbjct: 241 TVSSAIDKTIDKAKSL 256
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 124

A DNA sequence (GBSx0129) was identified in *S.agalactiae* <SEQ ID 421> which encodes the amino acid sequence <SEQ ID 422>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3405(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA56994 GB:X81089 glutamyl-aminopeptidase [Lactococcus lactis]
 Identities = 219/354 (61%), Positives = 273/354 (76%), Gaps = 1/354 (0%)

Query: 3   DLFNKIKTVTELDGIAGYEHNIRNFLRQEITPLVDQVETDGLGGIFGVKNTHETNAPKVM 62
           +LF+K+K +TE+   +G+E  +R++L+  +  L  Q E DGLGGIF  K +   NAP++M
Sbjct: 2   ELFDKVKALTEIQATSGFEGPVRDYLKARMVELGYQPEFDGLGGIFVTKASKVENAPRIM 61

Query: 63  VAAHMDEVGFMVSHIQPDGTFRVLEVGGWNPLVVSSQRFTLYTRSGDAIPVISGSVPPHF 122
```

```
            VAAHMDEVGFMVS I+ DGTFRV+ +GGWNPLVVS QRFTL+TR+G  IPV++G +PPH
Sbjct: 62  VAAHMDEVGFMVSSIKADGTFRVVPLGGWNPLVVSGQRFTLFTRTGKKIPVVTGGLPPHL 121

Query: 123 LRGQSGGTTLPKISDIVFDGGFTDKNEAESFGIAPGDIIVPKSETILTANQKHIMSKAWD 182
            LRG      +P ISDI+FDG F +  EA  FGIA GD+I+P++ETIL+AN K+I+SKAWD
Sbjct: 122 LRGTGVTPQIPAISDIIFDGAFENAAEAAEFGIAQGDLIIPETETILSANGKNIISKAWD 181

Query: 183 NRYGVLMVTELLKSLKDQSLSNTLIAGANVQEEVGLRGAHVSTTKFNPDIFLAVDCSPAG 242
            NRYG LM+ ELL+ L D+ L  TLI GANVQEEVGLRGA VSTTKFNPD+F AVDCSPA
Sbjct: 182 NRYGCLMILELLEFLADKELPVTLIIGANVQEEVGLRGAKVSTTKFNPDLFFAVDCSPAS 241

Query: 243 DIYG-EQGKIGEGTLIRFYDPGHIMLKDMRDFLLTTAEEAGIKYQYYAANGGTDAGAAHL 301
            D +G + G++GEGT +RF+DPGHIML  M++FLL TA  A +K Q Y A GGTDAGAAHL
Sbjct: 242 DTFGDDNGRLGEGTTLRFFDPGHIMLPGMKNFLLDTANHAKVKTQVYMAKGGTDAGAAHL 301

Query: 302 KNSGIPSTTIGVCARYIHSHQTLYAMDDFLQAQAYLQAIVNKLDRSTVDIIKGY 355
             N G+PSTTIGV ARYIHSHQT++ +DDFLQAQ +L+AI+  L+   V  IK Y
Sbjct: 302 ANGGVPSTTIGVVARYIHSHQTIFNIDDFLQAQTFLRAIITSLNTEKVAEIKNY 355
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 423> which encodes the amino acid sequence <SEQ ID 424>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2747(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 276/355 (77%), Positives = 322/355 (89%)

Query: 1   MSDLFNKIKTVTELDGIAGYEHNIRNFLRQEITPLVDQVETDGLGGIFGVKNTHETNAPK 60
            M+DLF+KIK VTELDGIAGYEH++R++LR +ITPLVD+VETDGLGGIFG++++    AP+
Sbjct: 1   MTDLFSKIKEVTELDGIAGYEHSVRDYLRTKITPLVDRVETDGLGGIFGIRDSKAEKAPR 60

Query: 61  VMVAAHMDEVGFMVSHIQPDGTFRVLEVGGWNPLVVSSQRFTLYTRSGDAIPVISGSVPP 120
            ++VAAHMDEVGFMVS I+ DGT RV+ +GGWNPLVVSSQRFTLYTR+G  IP+ISGSVPP
Sbjct: 61  ILVAAHMDEVGFMVSDIKVDGTLRVVGIGGWNPLVVSSQRFTLYTRTGQVIPLISGSVPP 120

Query: 121 HFLRGQSGGTTLPKISDIVFDGGFTDKNEAESFGIAPGDIIVPKSETILTANQKHIMSKA 180
            HFLRG +G  +LP I DIVFDGGFTDK EAE FGI PGDII+P+SETILTANQK+I+SKA
Sbjct: 121 HFLRGANGSASLPHIEDIVFDGGFTDKAEAERFGITPGDIIIPQSETILTANQKNIISKA 180

Query: 181 WDNRYGVLMVTELLKSLKDQSLSNTLIAGANVQEEVGLRGAHVSTTKFNPDIFLAVDCSP 240
            WDNRYGVLM+TE+L++LK Q L+NTLIAGANVQEEVGLRGAHVSTTKF+P++F AVDCSP
Sbjct: 181 WDNRYGVLMITEMLEALKGQDLNNTLIAGANVQEEVGLRGAHVSTTKFDPELFFAVDCSP 240

Query: 241 AGDIYGEQGKIGEGTLIRFYDPGHIMLKDMRDFLLTTAEEAGIKYQYYAANGGTDAGAAH 300
            AGDIYG  G IG+GTL+RFYDPGH+MLKDMRDFLLTTAEEAG+ +QYY   GGTDAGAAH
Sbjct: 241 AGDIYGNPGTIGDGTLLRFYDPGHVMLKDMRDFLLTTAEEAGVNFQYYCGKGGTDAGAAH 300

Query: 301 LKNSGIPSTTIGVCARYIHSHQTLYAMDDFLQAQAYLQAIVNKLDRSTVDIIKGY 355
            L+N G+PSTTIGVCARYIHSHQTLYAMDDF++AQA+LQAI+ KLDRSTVD+IK Y
Sbjct: 301 LQNGGVPSTTIGVCARYIHSHQTLYAMDDFVEAQAFLQAIIKKLDRSTVDLIKCY 355
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 125

A DNA sequence (GBSx0130) was identified in *S.agalactiae* <SEQ ID 425> which encodes the amino acid sequence <SEQ ID 426>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1672(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 126

A DNA sequence (GBSx0131) was identified in *S.agalactiae* <SEQ ID 427> which encodes the amino acid sequence <SEQ ID 428>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.28   Transmembrane   18 -  34 (  17 -  34)

----- Final Results -----
               bacterial membrane --- Certainty=0.1914(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 429> which encodes the amino acid sequence <SEQ ID 430>. Analysis of this protein sequence reveals the following:

```
     Possible site: 21
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -6.16   Transmembrane   12 -  28 (   8 -  30)

----- Final Results -----
               bacterial membrane --- Certainty=0.3463(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 30/91 (32%), Positives = 48/91 (51%)

Query: 13  MKNKKILFGTGLAGVGLLAAAGYTLTKKVTDYKRQQITQTLREFFSQMGDIQVFYFNEFE 72
           M  KKI   +G+ G  L    G  +     D +R+Q+T+ LR FFS +G I+V Y N  +
Sbjct: 4   MSKKKIGMISGIFGFSLAIGLGIVIKDYCQDRQRRQMTRDLRTFFSPLGQIEVLYINPCQ 63

Query: 73  SDIKMTSGGLVLEDGRIFEFIYRQGVLDYVE 103
                SGG+V+ +G+ ++F Y    + + E
Sbjct: 64  VKQDYISGGVVMSNGKQYQFTYHSRQISFEE 94
```

A related GBS gene <SEQ ID 8497> and protein <SEQ ID 8498> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 4
SRCFLG: 0
McG: Length of UR:   21
```

```
     Peak Value of UR:   2.30
     Net Charge of CR: 3
McG: Discrim Score:      6.28
GvH: Signal Score (-7.5): -1.46
     Possible site: 19
>>> Seems to have a cleavable N-term signal seq.
Amino Acid Composition: calculated from 20
ALOM program   count: 0 value:  22.60 threshold:  0.0
   PERIPHERAL  Likelihood = 22.60      29
 modified ALOM score:  -5.02

*** Reasoning Step: 3

Rule gpo1

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

SEQ ID 8498 (GBS214) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 40 (lane 3; MW 13.9kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 46 (lane 6; MW 39kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 127

A DNA sequence (GBSx0132) was identified in *S.agalactiae* <SEQ ID 431> which encodes the amino acid sequence <SEQ ID 432>. This protein is predicted to be thioredoxin H1 (trxA). Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2350(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB06972 GB:AP001518 thioredoxin H1 [Bacillus halodurans]
 Identities = 47/90 (52%), Positives = 66/90 (73%)

Query: 14  IDSTKKVVFFFTADWCPDCQFIYPVMPSIEKDFSDFVFVRVNRDDYIELAQQWNIFGIPS 73
           + + + VVF F+ADWCPDC+ I P +P +E+ + ++ F  VNRDD+IEL Q+ +IFGIPS
Sbjct: 13  VKNQENVVFLFSADWCPDCRVIEPFLPELEQTYDEYQFYYVNRDDFIELCQELDIFGIPS 72

Query: 74  FVVVENGQELGRLVNKNRKTKAEITKFLAE 103
           F+   NG+E  R V+K+RKTK EI +FL E
Sbjct: 73  FLFYSNGEERSRFVSKDRKTKEEIERFLTE 102
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 433> which encodes the amino acid sequence <SEQ ID 434>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1997(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 70/102 (68%), Positives = 81/102 (78%)

Query: 1   MILPESYEEIAAYIDSTKKVVFFFTADWCPDCQFIYPVMPSIEKDFSDFVFVRVNRDDYI 60
           MI P SYE +A  I+   K+V FFTADWCPDCQFIYP+MP IE + +D  FV VNRD +I
Sbjct: 1   MIRPTSYESLATLIEKEDKLVLFFTADWCPDCQFIYPIMPEIEAELTDMTFVCVNRDQFI 60

Query: 61  ELAQQWNIFGIPSFVVVENGQELGRLVNKNRKTKAEITKFLA 102
           E+AQ+WNIFGIPSFVV+E GQE+GRLVNK RKTK EI  FLA
Sbjct: 61  EVAQKWNIFGIPSFVVIEKGQEVGRLVNKMRKTKTEIMHFLA 102
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 128

A DNA sequence (GBSx0133) was identified in *S.agalactiae* <SEQ ID 435> which encodes the amino acid sequence <SEQ ID 436>. This protein is predicted to be phenylalanyl-tRNA synthetase beta subunit, non-spirochete. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1310(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC00291 GB:AF008220 YtpR [Bacillus subtilis]
 Identities = 78/196 (39%), Positives = 125/196 (62%), Gaps = 1/196 (0%)

Query: 5   YNREHVGDTLMVIVKDSQGAKLDVDRRGQVARVYLQDSKETVAWNIFEVSSLIVIEGAGQ 64
           YN+E VGDTL++ ++D    +L  ++ G V +++  ++KET  +NIF  SS + I+  G
Sbjct: 5   YNKEGVGDTLLISLQDVTREQLGYEKHGDVVKIFNNETKETTGFNIFNASSYLTIDENGP 64

Query: 65  ITLSDQDIKILNAELLKEGFEDSLVNNIEPTFVVAQIKEIIDHPDSDHLHICQAEINDGK 124
           + LS+  ++ +N  L + G E++LV ++ P FVV  ++     HP++D L +C+  + + +
Sbjct: 65  VALSETFVQDVNEILNRNGVEETLVVDLSPKFVVGYVESKEKHPNADKLSVCKVNVGE-E 123

Query: 125 TVQIVCGAPNASVGLKTVAALPGAMMPNGSLIFPGKLRGEDSFGMLCSARELALPNAPQV 184
           T+QIVCGAPN   G K V A  GA+MP+G +I   +LRG  S GM+CSA+EL LP+AP
Sbjct: 124 TLQIVCGAPNVDQGQKVVVAKVGAVMPSGLVIKDAELRGVPSSGMICSAKELDLPDAPAE 183

Query: 185 RGIIELSDQVIVGESF 200
           +GI+ L      G++F
Sbjct: 184 KGILVLEGDYEAGDAF 199
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 437> which encodes the amino acid sequence <SEQ ID 438>. Analysis of this protein sequence reveals the following:

```
     Possible site: 47
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.49   Transmembrane   90 - 106 (  90 - 107)

----- Final Results -----
               bacterial membrane --- Certainty=0.1595(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB06970 GB:AP001518 phenylalanyl-tRNA synthetase (beta subunit)
           [Bacillus halodurans]
 Identities = 84/196 (42%), Positives = 124/196 (62%), Gaps = 1/196 (0%)

Query: 5   YNKEQVGDVLMVILQDTKDIKRQVERKGKVARVFAEESGKTLAWNIFEASSLITIEGNGQ 64
           YN++ +GD +++++ + +   R  ER+G V R++   +GKT  +N+F AS      G G
Sbjct: 5   YNEKGIGDTILIVIDEVEPANRAYERQGDVVRIYHLGTGKTTGYNLFHASKYGEFNGQGL 64

Query: 65  IFLTDENLARLNAELAKEGFSERLEPIVGPVFVVGQIVEMVAHPDSDHLNICQVAIGEDQ 124
           + LTD  +A L     K G +  LE  + P FVVG +     HP++D L+IC+V +G D
Sbjct: 65  LELTDSLVATLEQAFQKNGVNWTLEVDLSPKFVVGFVQSKDKHPNADKLSICKVDVGSD- 123

Query: 125 TVQIVAGAPNAALGLKTIVALPGAIMPNGSLIFPGKLRGEESYGMMCSPRELALPNAPQK 184
           T+QIV GAPN   G K +VAL GA+MP+G +I P  LRG  S GM+CS +ELALP+AP++
Sbjct: 124 TLQIVCGAPNVEAGQKVVVALEGAVMPSGLVIKPTSLRGVSSTGMICSAKELALPDAPEE 183

Query: 185 RGIIEFDESAVVGEAF 200
           +GI+  D+S  VG +F
Sbjct: 184 KGILVLDDSYEVGTSF 199
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 133/207 (64%), Positives = 167/207 (80%)

Query: 1   MIFTYNREHVGDTLMVIVKDSQGAKLDVDRRGQVARVYLQDSKETVAWNIFEVSSLIVIE 60
           MIF YN+E VGD LMVI++D++  K  V+R+G+VARV+ ++S +T+AWNIFE SSLI IE
Sbjct: 1   MIFAYNKEQVGDVLMVILQDTKDIKRQVERKGKVARVFAEESGKTLAWNIFEASSLITIE 60

Query: 61  GAGQITLSDQDIKILNAELLKEGFEDSLVNNIEPTFVVAQIKEIIDHPDSDHLHICQAEI 120
           G GQI L+D+++  LNAEL KEGF + L   + P FVV QI E++ HPDSDHL+ICQ  I
Sbjct: 61  GNGQIFLTDENLARLNAELAKEGFSERLEPIVGPVFVVGQIVEMVAHPDSDHLNICQVAI 120

Query: 121 NDGKTVQIVCGAPNASVGLKTVAALPGAMMPNGSLIFPGKLRGEDSFGMLCSARELALPN 180
            + +TVQIV GAPNA++GLKT+ ALPGA+MPNGSLIFPGKLRGE+S+GM+CS RELALPN
Sbjct: 121 GEDQTVQIVAGAPNAALGLKTIVALPGAIMPNGSLIFPGKLRGEESYGMMCSPRELALPN 180

Query: 181 APQVRGIIELSDQVIVGESFDANKHWK 207
           APQ RGIIE  +  +VGE+FD  KHWK
Sbjct: 181 APQKRGIIEFDESAVVGEAFDPAKHWK 207
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 129

A DNA sequence (GBSx0135) was identified in *S.agalactiae* <SEQ ID 439> which encodes the amino acid sequence <SEQ ID 440>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3052(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB81904 GB:U92974 unknown [Lactococcus lactis]
 Identities = 69/241 (28%), Positives = 117/241 (47%), Gaps = 15/241 (6%)

Query: 7   YKEMLAKPWGKIQYEITFAQL--SHIKNQNVLDFGAGFCLTEQHLAKEN-NVTAIEPNPK 63
           Y E+  KPWG++ Y++ F QL  +  K+  +L FG+GF  TE  L ++   VT  EP+ +
Sbjct: 23  YAEVFEKPWGRMFYDLLFPQLLPNLTKDSKILSFGSGFGRTETFLEEQGFEVTGYEPDVE 82
```

```
Query: 64   LLYDNQSDNIYKILGSYEALRD-LPDQSFDTIICHNVLEYIDKHNHPAYFDEFSRLLKPN 122
             L         ++ G+++   + + ++ +D I+ HNVLEY+   +     +    LL
Sbjct: 83   KLEMMSDQTFRQLTGTFDDFAETVKNERYDVILIHNVLEYV--LDRKVVLELLLSLLTDG 140

Query: 123 GELSLIKHNITGKILQSVIFSNDTSTAMELLTGEANFKSASFDQGNIYT-----LEELKQ 177
           G LS++KH+  G +++     ++   A+++   EA    AS + G+I       L +
Sbjct: 141 GTLSIVKHSKYGSMIEMAAGRDNPQAALDVYENEA---VASHNHGDILVYDDDWLTDFVA 197

Query: 178 NTNLLVERYQGIRTFYSLQPN-HFKTETGWLNKMLAIELSVADKAPYKDIAFLQHITLKKS 237
           N  L ++   GIR FY +  N    K    W   ML +E  VA       +A L H+  KKS
Sbjct: 198 NYKLKLQEKFGIRHFYGISQNAEIKETENWYQPMLKLEQKVAKDQTLYPVARLHHLIFKKS 258
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 130

A DNA sequence (GBSx0136) was identified in *S.agalactiae* <SEQ ID 441> which encodes the amino acid sequence <SEQ ID 442>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3479(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF74079 GB:AF212845 putative single stranded binding protein
           [Lactococcus lactis bacteriophage ul36]
 Identities = 64/141 (45%), Positives = 92/141 (64%), Gaps = 10/141 (7%)

Query: 1   MYNKVIMIGRLTAKPEMVKTPTDKSVTRATVAVNRRFKGSNGEREADFINVVMWGRLAET 60
           M N V ++GR+T +PE+  TP +K+V   T+AVNR FK +NGEREADFI+ V+WG+ AE
Sbjct: 1   MINNVTLVGRITKEPELRYTPQNKAVATFTLAVNRAFKNANGEREADFISCVIWGKSAEN 60

Query: 61  LASYGTKGSLISIDGELRTRKYE-KDGQTHYITEVLASSFQLLESRAQ---------RAM 110
           LA++  KG LI + G ++TR YE + GQ  YITEV+AS+FQ+LE   Q          +
Sbjct: 61  LANWTHKGQLIGVIGNIQTRNYENQQGQRVYITEVVASNFQVLEKSNQANGERISNPASK 120

Query: 111 RENNVSGDLSDLVLEEEELPF 131
            +NN S     + + +++LPF
Sbjct: 121 PQNNDSFGSDPMEISDDDLPF 141
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 443> which encodes the amino acid sequence <SEQ ID 444>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1817(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 102/131 (77%), Positives = 116/131 (87%)

Query: 1   MYNKVIMIGRLTAKPEMVKTPTDKSVTRATVAVNRRFKGSNGEREADFINVVMWGRLAET 60
```

```
            MYNKVI IGRL AKPE+VKT TDK V R ++AVNRRFK ++GEREADFI+VV+WG+LAET
Sbjct: 1    MYNKVIAIGRLVAKPELVKTATDKHVARLSLAVNRRFKNASGEREADFISVVVWGKLAET 60

Query: 61   LASYGTKGSLISIDGELRTRKYEKDGQTHYITEVLASSFQLLESRAQRAMRENNVSGDLS 120
            L SY +KGSL+SIDGELRTRKY+KDGQ HY+TEVL  SFQLLESRAQRAMRENNV+ DL
Sbjct: 61   LVSYASKGSLMSIDGELRTRKYDKDGQVHYVTEVLCQSFQLLESRAQRAMRENNVTNDLV 120

Query: 121  DLVLEEEELPF 131
            DLVLEE+ LPF
Sbjct: 121  DLVLEEDTLPF 131
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 131

A DNA sequence (GBSx0137) was identified in *S.agalactiae* <SEQ ID 445> which encodes the amino acid sequence <SEQ ID 446>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2235(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9493> which encodes amino acid sequence <SEQ ID 9494> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAC13072 GB:AL445503 putative hydrolase [Streptomyces
            coelicolor]
 Identities = 63/179 (35%), Positives = 91/179 (50%), Gaps = 2/179 (1%)

Query: 33   IIFDMDGVIVDSEYTFLDNKTEMLREEGI-DTDVSYQYQYMGTTFEFMWQAMKEEFGLPK 91
            +IFD+DG +VDSE  + +     L E G+ D   +    Y+G + +      K  +GL
Sbjct: 12   VIFDLDGTLVDSEPHYYEAGRRTLAEYGVPDFSWADHEAYVGISTQETVADWKRRYGLRA 71

Query: 92   TVKEYIAEMNRRRQAIVARDGVRPIKGAQRLIHWLHQHGYRLAVASSSPMVDIKRNLKEL 151
            TV+E +A  NR     + AR   R      ++ +  L   G  +AVAS S    I   L
Sbjct: 72   TVEELLAVKNRHYLGL-ARTSARAYPEMRKFVELLAGEGVPMAVASGSSPEAIAAILART 130

Query: 152  GVTECFEYMVTGEDVSSSKPAPDVFLRAAELLDVDPKVCIVIEDTRNGSLAAKAAGMYC 210
            G+      +V+ ++V+  KPAPDVFL AA  L  +P  C+V+ED   G+ AA AAGM C
Sbjct: 131  GLDAHLRTVVSADEVARGKPAPDVFLEAARRLGTEPARCVVLEDAAPGAAAAHAAGMRC 189
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 447> which encodes the amino acid sequence <SEQ ID 448>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3706(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 62/202 (30%), Positives = 100/202 (48%), Gaps = 1/202 (0%)

Query: 29   MEKVIIFDMDGVIVDSEYTFLDNKTEMLREEGIDTDVSYQYQYMGTTFEFMWQAMKEEFG 88
```

```
            M K IIFDMDGV+ D+E  +L  + +  + +GI  D      ++G   + +W+ +  +
Sbjct: 3    MIKGIIFDMDGVLFDTEPFYLRRREDFFKTKGIPIDHLNSKDFIGGNLQELWKELLGKNR 62

Query: 89   LPKTVKEYIAEMNRRRQAIVARDGVRPIKGAQRLIHWLHQHGYRLAVASSSPMVDIKRNL 148
             5                VK    + +  +QA         I      +  L + G +LAVAS+S   D+   L
Sbjct: 63   DDAIVKAITTDYDAYKQAHKPPYQKLLITEVNSCLEQLEKQGIKLAVASNSKRQDVLLAL 122

Query: 149  KELGVTECFEYMVTGEDVSSSKPAPDVFLRAAELLDVDPKVCIVIEDTRNGSLAAKAAGM 208
             +   + + FE ++  EDVS  KP PD++ +A + L +  K  +V+ED++ G  AAKAA +
Sbjct: 123  ETTQIKDYFEIILAREDVSRGKPYPDIYNKAVQKLGLQKKQLLVVEDSQKGIAAAKAANL 182

Query: 209  YCFGFANPDYPPQDLSMADKVI 230
               F   +  Y   D S AD  I
Sbjct: 183  TVFAITDYRY-GIDQSQADHKI 203
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 132

A DNA sequence (GBSx0138) was identified in *S.agalactiae* <SEQ ID 449> which encodes the amino acid sequence <SEQ ID 450>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.22   Transmembrane   16 -  32 (  16 -  32)

----- Final Results -----
               bacterial membrane --- Certainty=0.1086(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 133

A DNA sequence (GBSx0139) was identified in *S.agalactiae* <SEQ ID 451> which encodes the amino acid sequence <SEQ ID 452>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -5.04   Transmembrane   28 -  44 (  27 -  45)

----- Final Results -----
               bacterial membrane --- Certainty=0.3017(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 134

A DNA sequence (GBSx0140) was identified in *S.agalactiae* <SEQ ID 453> which encodes the amino acid sequence <SEQ ID 454>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.72   Transmembrane   38 -  54 (  34 -  60)
   INTEGRAL    Likelihood = -7.70   Transmembrane    4 -  20 (   1 -  22)
   INTEGRAL    Likelihood = -4.99   Transmembrane  153 - 169 ( 150 - 171)
   INTEGRAL    Likelihood = -2.55   Transmembrane  179 - 195 ( 178 - 198)
   INTEGRAL    Likelihood = -2.39   Transmembrane   93 - 109 (  93 - 109)
   INTEGRAL    Likelihood = -1.17   Transmembrane  116 - 132 ( 116 - 133)
   INTEGRAL    Likelihood = -0.43   Transmembrane  344 - 360 ( 344 - 360)

----- Final Results -----
               bacterial membrane --- Certainty=0.5288(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB14853 GB:Z99118 two-component sensor histidine kinase
          [Bacillus subtilis]
 Identities = 254/585 (43%), Positives = 371/585 (63%), Gaps = 9/585 (1%)

Query: 2   LMVLLFQRLGIIMILAFLLVNNSYFRQLIEERSK-RETVVLVIIFGLFVIISNITGIEIK 60
           LM+++ +R+GII+IL F+L +   FRQ ++ +   +   +L+ IF LF IISN TGIEI+
Sbjct: 4   LMIMMLERVGIIVILGFILAHTKLFRQALQNQDGYKGKAILISIFSLFSIISNYTGIEIQ 63

Query: 61  GDRSLVERPFLTTISHSDSLANTRTLVITTASLVGGPLVGSIVGFIGGVHRFFQGSFSGS 120
            +  +V   ++ TI  S S+ANTR L +    L+GGP VG+ +G + G+HRF  G  +
Sbjct: 64  RNM-IVNNDWVFTIDPSGSIANTRILGVEIGGLLGGPFVGAGIGILAGLHRFSLGGSTAL 122

Query: 121 FYIVSSVLVGIVSGKIGDKLKENHLYPSTSQVILISIIAESIQMLFVGIFT-----GWEL 175
              VSS+L G+++G IG    + +  P+     L+ I  ES+QM+ + +         WEL
Sbjct: 123 SCAVSSILAGVLAGLIGRYFTKRYRMPTPRIAALVGIGMESLQMIIILLMAKPFSDAWEL 182

Query: 176 VKMIVIPMMILNSLGSTLFLAILKTYLSNESQLRAVQTRDVLELTRQTLPYLRQGLTPQS 235
           V MI IPM+++N  GS +FL+I++  + E Q RA++T  VL +  QTLP+ RQGL   S
Sbjct: 183 VSMIGIPMILINGTGSFIFLSIIQAIIRKEEQARALETHRVLTIADQTLPFFRQGLNENS 242

Query: 236 ARSVCEIIKRHTNFDAVGLTDRSNVLAHIGVGHDHHIAGQPVKTDLSKSVIFDGEPRIAQ 295
            +SV  II + T  DAV LTD+  +LAH+G G DHHI  + + T LSK VI  G    A
Sbjct: 243 CKSVAAIIHKLTGTDAVSLTDKEKILAHVGAGMDHHIPSKSLITGLSKKVIKTGHIMKAI 302

Query: 296 DKAAISCPDHNCQLNSAIVVPLKINDKTVGALKMYFAGDKTMSEVEENLVLGLAQIFSGQ 355
            +  I C    C L++AIV+PL  N  T+G LKMYF     +S+VEE L  GLA +FS Q
Sbjct: 303 SQEEIECTHAECPLHAAIVLPLTSNGNTIGTLKMYFKSPAGLSQVEEELAEGLAMLFSTQ 362

Query: 356 LAMGITEEQNKLASMAEIKALQAQINPHFFFNAINTISALIRIDSDKARYALMQLSTFFR 415
           L +G  E Q+KL   AEIKALQAQ+NPHF FNAINTISAL R D +K R  L+QLS +FR
Sbjct: 363 LELGEAELQSKLLKDAEIKALQAQVNPHFLFNAINTISALCRTDVEKTRKLLLQLSVYFR 422

Query: 416 TSLQGGQDREVTLEQEKSHVDAYMNVEKLRFPDKYQLSYDI-SAPEKMKLPPFGLQVLVE 474
           ++LQG +   + L +E +H++AY+++E+ RFP KY++  +I S  E++++PPF LQVLVE
Sbjct: 423 SNLQGARQLLIPLSKELNHLNAYLSLEQARFPGKYKIELNIDSRLEQIEIPPFVLQVLVE 482

Query: 475 NAVRHAFKERKTDNHILVQIKPDGHYYCVSVSDNGQGISDTIIDKLGQETVAESKGTGTA 534
           NA+RHAF +++    + V +  D     + V+DNG+GI   ++ +LG++     +GTGTA
Sbjct: 483 NALRHAFPKKQDICKVTVCVLSDDASVYMKVADNGRGIPPDVLPELGKKPFPSKEGTGTA 542

Query: 535 LVNLNNRLNLLYGSVSCLHFSSD-KNGTKVWYRIPNRIREDEHEN 578
           L NLN RL  L+G  + LH SS+   GT+V +++P +  ++  E+
Sbjct: 543 LYNLNQRLIGLFGQQAALHISSEVHKGTEVSFQVPMQQMKEGEEH 587
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 455> which encodes the amino acid sequence <SEQ ID 456>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1771(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 75/245 (30%), Positives = 117/245 (47%), Gaps = 22/245 (8%)

Query: 348 LAQIFSGQL-----AMGITEEQNKLASMAEIKALQAQINPHFFFNAINTISALIRI-DSD 401
           LAQ F+  L      M    ++ K      ++AL +QINPHF +N ++TI  +    DS
Sbjct: 4   LAQQFNALLDQIDSLMVAVADKEKAIGQYRLQALASQINPHFLYNTLDTIIWMAEFNDSK 63

Query: 402 KARYALMQLSTFFRTSLQGGQDREVTLEQEKSHVDAYMNVEKLRFPDKYQLSYDISAPE- 460
           +       L+ +FR +L  G +  + L  E  HV  Y+ ++K R+ DK  LSY++   +
Sbjct: 64  RVVEVTKSLAKYFRLALNQGNEY-IRLADELDHVSQYLFIQKQRYGDK--LSYEVQGLDV 120

Query: 461 --KMKLPPFGLQVLVENAVRHAFKERKTDNHILVQIKPDGHYYCVSVSDNGQGISDTIID 518
               +P   LQ LVENA+ H  KE      I V +     +  ++V DNG+GI D+ +
Sbjct: 121 YADFVIPKLILQPLVENAIYHGIKEVDRKGMIKVTVSDTAQHLMLTVWDNGKGIEDSSLT 180

Query: 519 KLGQETVAESKGTGTALVNLNNRLNLLYGS--VSCLHFSSDKNGTKVWYRIPNR---IRE 573
              Q  +A     G  L N++ RL L YG      +H  SD+  T++   +P     + +
Sbjct: 181 N-SQSLLARG---GVGLKNVDQRLKLHYGEGYHMTIHSQSDQ-FTEIQLSLPKMHELMAD 235

Query: 574 DEHEN 578
           D  EN
Sbjct: 236 DTQEN 240
```

SEQ ID 454 (GBS248d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 124 (lane 2-4; MW 71kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 124 (lane 5-7; MW 46kDa) and in Figure 180 (lane 2; MW 46kDa).

GBS248d-His was purified as shown in Figure 234, lane 3-4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 135

A DNA sequence (GBSx0141) was identified in *S.agalactiae* <SEQ ID 457> which encodes the amino acid sequence <SEQ ID 458>. This protein is predicted to be two-component response regulator (lytT). Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3230(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9495> which encodes amino acid sequence <SEQ ID 9496> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB14852 GB:Z99118 two-component response regulator [Bacillus subtilis]
 Identities = 105/244 (43%), Positives = 157/244 (64%), Gaps = 6/244 (2%)

Query: 3   MKILILDDEMFARQELSFLVEHSQEVDNPEIFQAEDISEAEKILFRQQIDLIFLDISLSE 62
           +++LI+DDEM AR EL++L++ +   D  EI +AE+I  A   +  Q+ DL+FLD+ LS
Sbjct: 2   LRVLIVDDEMLARDELAYLLKRTN--DEMEINEAENIESAFDQMMDQKPDLLFLDVDLSG 59

Query: 63  ENGFTLANQLSQLAHPPLVVFATAYDNYAVKAFESNAVDYIMKPFEQQRVDMALSKVKKL 122
           ENGF +A +L ++ HPP +VFATAYD YA+KAFE +A+DY+ KPF+++R+   L K KK+
Sbjct: 60  ENGFDIAKRLKKMKHPPAIVFATAYDQYALKAFEVDALDYLTKPFDEERIQQTLKKYKKV 119

Query: 123 SQLTTASDVEQAIPKKASVELLTLTLSDRSVVVKMQDIVAASVEDGELTVSTVQKTYTIR 182
           ++      VE      A    L L++ +  V+V  +DI+ A  EDG + V T    +YT+
Sbjct: 120 NR----DIVETEQNSHAGQHKLALSVGESIVIVDTKDIIYAGTEDGHVNVKTFDHSYTVS 175

Query: 183 KTLNWFKSRAVAPYFLQIHRNTVINLEMIEEIQPWFNHTLLLIMSNGEKFPVGRSYLKDL 242
            TL   + +     F+++HR+ V+N E I+EIQPWFN T  LIM +G K PV R+Y K+L
Sbjct: 176 DTLVVIEKKLPDSDFIRVHRSFVVNTEYIKEIQPWFNSTYNLIMKDGSKIPVSRTYAKEL 235

Query: 243 NEHL 246
            + L
Sbjct: 236 KKLL 239
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 459> which encodes the amino acid sequence <SEQ ID 460>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3818(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 44/148 (29%), Positives = 84/148 (56%), Gaps = 5/148 (3%)

Query: 5   ILILDDEMFARQELSFLVEHSQ-EVDNPEIFQAEDISEAEKILFRQQIDLIFLDISLSEE 63
           +LI++DE   RQ +  LV+ SQ ++D   + +AE+   A  +  ++  D++  DI++ +
Sbjct: 4   LLIVEDEYLVRQGIRSLVDFSQFKIDR--VNEAENGQLAWDLFQKEPYDIVLTDINMPKL 61

Query: 64  NGFTLANQLSQLAHPPLVVFATAYD--NYAVKAFESNAVDYIMKPFEQQRVDMALSKVKK 121
           NG  LA  + Q +    +VF T YD  NYA+ A +  A DY++KPF +  V+  L K++K
Sbjct: 62  NGIQLAELIKQESPQTHLVFLTGYDDFNYALSALKLGADDYLLKPFSKADVEDMLGKLRK 121

Query: 122 LSQLTTASDVEQAIPKKASVELLTLTLS 149
             +L+  ++  Q + ++   E+  + ++
Sbjct: 122 KLELSKKTETIQELVEQPQKEVSAIAMA 149
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 136

A DNA sequence (GBSx0142) was identified in *S.agalactiae* <SEQ ID 461> which encodes the amino acid sequence <SEQ ID 462>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.0266(Affirmative) < succ>
```

```
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 137

A DNA sequence (GBSx0143) was identified in *S.agalactiae* <SEQ ID 463> which encodes the amino acid sequence <SEQ ID 464>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.89   Transmembrane  104 - 120 (  99 - 134)
   INTEGRAL    Likelihood = -5.89   Transmembrane   47 -  63 (  46 -  65)
   INTEGRAL    Likelihood = -3.29   Transmembrane   22 -  38 (  21 -  39)
   INTEGRAL    Likelihood = -2.81   Transmembrane   74 -  90 (  70 -  92)

----- Final Results -----
               bacterial membrane --- Certainty=0.5755(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8499> which encodes amino acid sequence <SEQ ID 8500> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB14851  GB:Z99118  similar  to  hypothetical  proteins  from  B.  subtilis  [Bacillus
subtilis]
 Identities = 50/110 (45%), Positives = 82/110 (74%), Gaps = 2/110 (1%)

Query: 20  QMSIYAAILLVSQMISMLLPKSLPIPTTVIGLVLMYVLLTAKIIKVEWVDSFGALMISMI 79
           Q  I+A I+LVS MI+ ++P  +PIP +V+GLVL+++LL  K+IK+E V++ G  + S+I
Sbjct: 12  QAFIFAVIMLVSNMIAAIVP--IPIPASVVGLVLLFLLLCLKVIKLEQVETLGTSLTSLI 69

Query: 80  GFMFVPSGISVAANLDILKAEGLQLVAVITISTVVMLVVVAYVARLILAI 129
           GF+FVPSGISV  +L +++  GLQ+V VI ++T+++L      ++LIL++
Sbjct: 70  GFLFVPSGISVMNSLGVMQQYGLQIVLVILLATIILLGATGLFSQLILSL 119
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 138

A DNA sequence (GBSx0144) was identified in *S.agalactiae* <SEQ ID 465> which encodes the amino acid sequence <SEQ ID 466>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-12.21   Transmembrane  219 - 235 ( 208 - 241)
   INTEGRAL    Likelihood =-11.94   Transmembrane  103 - 119 (  99 - 133)
   INTEGRAL    Likelihood = -5.57   Transmembrane  157 - 173 ( 154 - 175)
   INTEGRAL    Likelihood = -1.70   Transmembrane   73 -  89 (  73 -  89)

----- Final Results -----
```

```
bacterial membrane --- Certainty=0.5883(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB14850 GB:Z99118 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 120/240 (50%), Positives = 159/240 (66%), Gaps = 10/240 (4%)

Query: 1   MELLKTPIFGICFSLILYTIGEHLFKKSKGFFLLQPLFFAMVSGIVILWLMSKGLGTDVK 60
            ME   +P FGI  SL  + IG  LFKK+KGFFL  PLF AMV GI  L          +
Sbjct: 1   MESTMSPYFGIVVSLAAFGIGTFLFKKTKGFFLFTPLFVAMVLGIAFL---------KIG 51

Query: 61  TFYTQAYKPGGDLIFWFLNPATIAFAVPLYKKNDVVKKYWVEILSSLVIGMIVSLILIVA 120
             F    Y  GG++I +FL PATIAFA+PLYK+ D +KKYW +I++S++ G I S+ ++
Sbjct: 52  GFSYADYNNGGEIIKFFLEPATIAFAIPLYKQRDKLKKYWWQIMASIIAGSICSVTIVYL 111

Query: 121 ISKMVGLSQVGIASMLPQAATTAIALPITAAIGGNTAVTAMACILNAVIIYALGKKLVSF 180
            ++K + L    + SMLPQAATTAIALP++  IGG + +TA A I NAVI+YALG   +
Sbjct: 112 LAKGIHLDSAVMKSMLPQAATTAIALPLSKGIGGISDITAFAVIFNAVIVYALGALFLKV 171

Query: 181 FHLNDSKIGAGLGLGTSGHTVGAAFALELGELQGAMAAIAVVVIGLVVDLVIPIFSHLIG 240
            F +  + I  GL LGTSGH +G A  +E+GE++ AMA+IAVVV+G+V  LVIP+F  LIG
Sbjct: 172 FKVK-NPISKGLALGTSGHALGVAVGIEMGEVEAAMASIAVVVVGVVTVLVIPVFVQLIG 230
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 139

A DNA sequence (GBSx0145) was identified in *S.agalactiae* <SEQ ID 467> which encodes the amino acid sequence <SEQ ID 468>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> May be a lipoprotein

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 508/542 (93%), Positives = 523/542 (95%)

Query: 1   MTKYLKYISFVALFLASIFLVACQNQNSQTKERTRKQRPKDELVVSMGAKLPHEFDPKDR 60
            ++KYLKY S + LFL  + LVACQ Q  QTKER RKQRPKDELVVSMGAKLPHEFDPKDR
Sbjct: 3   VSKYLKYFSIITLFLTGLILVACQQQKPQTKERQRKQRPKDELVVSMGAKLPHEFDPKDR 62

Query: 61  YGIHNEGNITHSTLLKRSPELDIKGELAKKYKISKDGLTWSFDLNDDFKFSNGEPVTADD 120
            YG+HNEGNITHSTLLKRSPELDIKGELAK Y +S+DGLTWSFDL+DDFKFSNGEPVTADD
Sbjct: 63  YGVHNEGNITHSTLLKRSPELDIKGELAKTYHLSEDGLTWSFDLHDDFKFSNGEPVTADD 122

Query: 121 VKFTYDMLKADGKAWDLTFIKNVEVVGKNQVNIHLTEAHSTFTAQLTEIPIVPKKHYNDK 180
            VKFTYDMLKADGKAWDLTFIKNVEVVGKNQVNIHLTEAHSTFTAQLTEIPIVPKKHYNDK
Sbjct: 123 VKFTYDMLKADGKAWDLTFIKNVEVVGKNQVNIHLTEAHSTFTAQLTEIPIVPKKHYNDK 182

Query: 181 YKSNPIGSGPYMVKEYKAGEQAIFVRNPYWHGKKPYFKKWTWVLLDENTALAALESGDVD 240
            YKSNPIGSGPYMVKEYKAGEQAIFVRNPYWHGKKPYFKKWTWVLLDENTALAALESGDVD
Sbjct: 183 YKSNPIGSGPYMVKEYKAGEQAIFVRNPYWHGKKPYFKKWTWVLLDENTALAALESGDVD 242

Query: 241 MIYATPELASKKVKGTRLLDIASNDVRGLSLPYVKKGVVKNSPDGYPVGNDVTSDPAIRK 300
            MIYATPELA KKVKGTRLLDI SNDVRGLSLPYVKKGV+ +SPDGYPVGNDVTSDPAIRK
```

```
Sbjct: 243 MIYATPELADKKVKGTRLLDIPSNDVRGLSLPYVKKGVITDSPDGYPVGNDVTSDPAIRK 302

Query: 301 ALTIGLNRQKVLDTVLNGYGKPAYSIIDRTPFWNPKTAIKDNKVAKAKQLLTKAGWKEQA 360
           ALTIGLNRQKVLDTVLNGYGKPAYSIID+TPFWNPKTAIKDNKVAKAKQLLTKAGWKEQA
Sbjct: 303 ALTIGLNRQKVLDTVLNGYGKPAYSIIDKTPFWNPKTAIKDNKVAKAKQLLTKAGWKEQA 362

Query: 361 DGSRKKGNLKSEFDLYYPTNDQLRANLAVEVAEQAKALGITIKLKASNWDEMATKSHDSA 420
           DGSRKKG+L + FDLYYPTNDQLRANLAVEVAEQAKALGITIKLKASNWDEMATKSHDSA
Sbjct: 363 DGSRKKGDLDAAFDLYYPTNDQLRANLAVEVAEQAKALGITIKLKASNWDEMATKSHDSA 422

Query: 421 LLYAGGRHHAQQFYESHYPSLAGKGWTNITFYNNPTVTKYLDKAMTSPDLDKANKYWKLA 480
           LLYAGGRHHAQQFYESH+PSLAGKGWTNITFYNNPTVTKYLDKAMTS DLDKAN+YWKLA
Sbjct: 423 LLYAGGRHHAQQFYESHHPSLAGKGWTNITFYNNPTVTKYLDKAMTSSDLDKANEYWKLA 482

Query: 481 QWDGKTGASTLGDLPNVWLVSLNHTYIGDKRINVGKQGVHSHGHDWSLLTNIAEWTWDES 540
           QWDGKTGASTLGDLPNVWLVSLNHTYIGDKRINVGKQGVHSHGHDWSLLTNIAEWTWDES
Sbjct: 483 QWDGKTGASTLGDLPNVWLVSLNHTYIGDKRINVGKQGVHSHGHDWSLLTNIAEWTWDES 542

Query: 541 AK 542
            K
Sbjct: 543 TK 544
```

There is also homology to SEQ ID 60.

A related GBS gene <SEQ ID 8501> and protein <SEQ ID 8502> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: 22   Crend: 5
McG: Discrim Score:     10.46
GvH: Signal Score (-7.5): -1.29
     Possible site: 22
>>> May be a lipoprotein
ALOM program   count: 0 value:   7.27 threshold:  0.0
   PERIPHERAL  Likelihood =  7.27     386
 modified ALOM score:  -1.95

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

SEQ ID 8502 (GBS106) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 18 (lane 3; MW 61kDa).

The GBS106-His fusion product was purified (Figure 194, lane 2) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 255A), FACS (Figure 255B), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 140

A DNA sequence (GBSx0146) was identified in *S.agalactiae* <SEQ ID 469> which encodes the amino acid sequence <SEQ ID 470>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4862(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 141

A DNA sequence (GBSx0147) was identified in *S.agalactiae* <SEQ ID 471> which encodes the amino acid sequence <SEQ ID 472>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.27   Transmembrane  252 - 268 ( 249 - 275)
   INTEGRAL    Likelihood = -5.73   Transmembrane   67 -  83 (  62 -  90)
   INTEGRAL    Likelihood = -5.26   Transmembrane  107 - 123 ( 104 - 134)
   INTEGRAL    Likelihood = -3.77   Transmembrane  153 - 169 ( 152 - 170)

----- Final Results -----
              bacterial membrane --- Certainty=0.3909(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9295> which encodes amino acid sequence <SEQ ID 9296> was also identified.

The protein differs from U78968 at the N-terminus:

```
Query: 1   MASVNYDTSLTPVQYKAIAHHYGLDKPAPVQYFIWLKNFIQGHLGTSLVYRQPVIDIIRS 60
           MASVNYDTSLTP QYKAIAHHYGLDKPA VQYFIWLKN IQG LGTSLVYRQPV DIIRS
Sbjct: 39  MASVNYDTSLTPAQYKAIAHHYGLDKPALVQYFIWLKNVIQGDLGTSLVYRQPVSDIIRS 98
```

There is also homology to SEQ ID 64.

A related GBS gene <SEQ ID 8471> and protein <SEQ ID 8472> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:      3.72
GvH: Signal Score (-7.5): -5.37
     Possible site: 40
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 5 value:  -7.27 threshold:  0.0
   INTEGRAL    Likelihood = -7.27   Transmembrane  290 - 306 ( 287 - 313)
   INTEGRAL    Likelihood = -5.89   Transmembrane   12 -  28 (  11 -  33)
   INTEGRAL    Likelihood = -5.73   Transmembrane  105 - 121 ( 100 - 128)
   INTEGRAL    Likelihood = -5.26   Transmembrane  145 - 161 ( 142 - 172)
   INTEGRAL    Likelihood = -3.77   Transmembrane  191 - 207 ( 190 - 208)
   PERIPHERAL  Likelihood =  2.97     245
 modified ALOM score:   1.95

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.3909(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

SEQ ID 8472 (GBS436) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 173 (lane 9; MW 54kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 142

A DNA sequence (GBSx0148) was identified in *S.agalactiae* <SEQ ID 473> which encodes the amino acid sequence <SEQ ID 474>. This protein is predicted to be transmembrane transport protein DppC (oppC). Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -8.28   Transmembrane   77 -  93 (  68 - 101)
   INTEGRAL    Likelihood = -7.80   Transmembrane  182 - 198 ( 180 - 204)
   INTEGRAL    Likelihood = -7.06   Transmembrane  112 - 128 ( 104 - 132)
   INTEGRAL    Likelihood = -5.10   Transmembrane  239 - 255 ( 235 - 258)

----- Final Results -----
               bacterial membrane --- Certainty=0.4312(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

There is homology to SEQ ID 68.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 143

A DNA sequence (GBSx0149) was identified in *S.agalactiae* <SEQ ID 475> which encodes the amino acid sequence <SEQ ID 476>. This protein is predicted to be ATPase protein DppD. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1957(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein differs from U78968 at the C-terminus:

```
Query: 241 QTEFARSLWRSLPQQEFLKGVTHDLRG 267
           QTEFAR LWR+LPQQ+FLKGVTHDLRG
Sbjct: 241 QTEFARRLWRTLPQQDFLKGVTHDLRG 267
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 477> which encodes the amino acid sequence <SEQ ID 478>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1957(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 255/267 (95%), Positives = 262/267 (97%)

Query: 1   MTETLLSIKDLSITFTQYGRFLKPFQSTPIQALNLEIKKGELLAIIGASGSGKSLLAHAI 60
           MTETLLSIKDLSITFTQYGRFLKPFQSTPIQALNLE+KKGELLAIIGASGSGKSLLAHAI
Sbjct: 1   MTETLLSIKDLSITFTQYGRFLKPFQSTPIQALNLEVKKGELLAIIGASGSGKSLLAHAI 60

Query: 61  MDILPKNASVTGDMIYRGQSLNSKRIKQLRGKDITLIPQSVNYLDPSTKVKHQVRLGISE 120
           MDILPKNA+VTGDMIYRGQSL SKRIKQLRGK++TLIPQSVNYLDPS KVKHQVRLGISE
Sbjct: 61  MDILPKNAAVTGDMIYRGQSLTSKRIKQLRGKEMTLIPQSVNYLDPSMKVKHQVRLGISE 120

Query: 121 NSKATQEGLFQQFGLKESDGDLYPFQLSGGMLRRVLFTTCISDKVSLIIADEPTPGLHPD 180
           N+KATQEGLFQQFGLKESDGDLYPFQLSGGMLRRVLFTTCISD VSLIIADEPTPGLHPD
Sbjct: 121 NAKATQEGLFQQFGLKESDGDLYPFQLSGGMLRRVLFTTCISDTVSLIIADEPTPGLHPD 180

Query: 181 ALQMVLDQLRSFADKGISVIFITHDIVAASQIADRITIFKEGKAIETAPASFFSGNGEQL 240
           ALQMVLDQLRSFADKGISVIFITHDIVAASQIADRITIFKEGKAIETAPASFFSG GEQL
Sbjct: 181 ALQMVLDQLRSFADKGISVIFITHDIVAASQIADRITIFKEGKAIETAPASFFSGGGEQL 240

Query: 241 QTEFARSLWRSLPQQEFLKGVTHDLRG 267
           QTEFAR LWR+LPQQ+FLKGVTHDLRG
Sbjct: 241 QTEFARRLWRTLPQQDFLKGVTHDLRG 267
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 144

A DNA sequence (GBSx0150) was identified in *S.agalactiae* <SEQ ID 479> which encodes the amino acid sequence <SEQ ID 480>. This protein is predicted to be ATPase protein DppE. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3783(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 481> which encodes the amino acid sequence <SEQ ID 482>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3383(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 188/205 (91%), Positives = 197/205 (95%)

Query: 1   MTLEAKKLGFYHKKDQWLFKEINLEVAPGQVLGIFGQSGCGKTSLSRVLAGFLHPKSGEV 60
           MTLEAKKLGFYHKKDQWLFKEI+LEVAPGQ+LGIFGQSGCGKTSLSRVLAGFL PKSGEV
Sbjct: 1   MTLEAKKLGFYHKKDQWLFKEIDLEVAPGQILGIFGQSGCGKTSLSRVLAGFLQPKSGEV 60

Query: 61  LVDGSNLPSKAFRPVQLIQQHPEKTMNPLWPMKKSLEEAYYPSRDLLDAFGIQEKWLNRR 120
           LVDGS+LP+KAFRPVQLIQQHPE+TMNPLWPMKKSLEEAYYPS+DL DAFGIQEKWL RR
Sbjct: 61  LVDGSHLPNKAFRPVQLIQQHPEQTMNPLWPMKKSLEEAYYPSQDLRDAFGIQEKWLKRR 120
```

```
Query: 121 PSELSGGELQRFSIVRSLHPETKYLIADEMTTMLDSITQASVWKSLLEIVKDRNLGLIVI 180
           PSELSGGELQRFSIVRSLHPETKYLIADEMTTMLDSITQASVWKSLLEIVKDRNLGLI+I
Sbjct: 121 PSELSGGELQRFSIVRSLHPETKYLIADEMTTMLDSITQASVWKSLLEIVKDRNLGLIII 180

Query: 181 SHDFAMLEKLCNQCYMIEENRIVSF 205
           SH+F MLEKLC+ CYMIEENR   F
Sbjct: 181 SHEFDMLEKLCDACYMIEENRTQLF 205
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 145

A DNA sequence (GBSx0151) was identified in *S.agalactiae* <SEQ ID 483> which encodes the amino acid sequence <SEQ ID 484>. This protein is predicted to be PTS system, trehalose-specific IIBC component (treB). Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.14   Transmembrane  468 - 484 ( 462 - 489)
   INTEGRAL    Likelihood = -8.23   Transmembrane  279 - 295 ( 275 - 306)
   INTEGRAL    Likelihood = -6.05   Transmembrane  112 - 128 ( 105 - 130)
   INTEGRAL    Likelihood = -3.35   Transmembrane  204 - 220 ( 203 - 222)
   INTEGRAL    Likelihood = -1.75   Transmembrane  255 - 271 ( 255 - 271)
   INTEGRAL    Likelihood = -1.54   Transmembrane  327 - 343 ( 326 - 344)
   INTEGRAL    Likelihood = -0.37   Transmembrane  422 - 438 ( 422 - 438)
   INTEGRAL    Likelihood = -0.06   Transmembrane  304 - 320 ( 304 - 320)

----- Final Results -----
              bacterial membrane --- Certainty=0.5055(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF94072 GB:AE004175 PTS system, trehalose-specific IIBC
           component [Vibrio cholerae]
 Identities = 225/484 (46%), Positives = 318/484 (65%), Gaps = 28/484 (5%)

Query: 5   KHDAKALLEAIGGKENISAVTHCATRMRFVLNDSSKAKVKVIEELPSVKGTFTNAGQFQV 64
           K D   L+E +GG+ NI++VTHC TR+RFVLN   +A    +E L  VKG FTNAGQFQV
Sbjct: 10  KQDVTRLIELVGGESNIASVTHCLTRLRFVLNQPEQADKAGLEALSMVKGCFTNAGQFQV 69

Query: 65  IIGNDVPIFYNAFVAVSGIEGVSKEAAKSAAQKNQNPLQRVLTMLAEIFTPIIPAIIVGG 124
           +IG +V   Y   +  +G + VSK+ AK AA++N N L+R ++ LAEIF P++PAII GG
Sbjct: 70  VIGTEVDQVYKMLLEQTGKQAVSKDDAKVAARQNMNVLERGISHLAEIFVPLLPAIITGG 129

Query: 125 LILGFRNILDAVPFEFLGQKVVDGVRQVDSSGHPIWNTLVDVSTFWSGVDSFLWLPGEAI 184
           LILGFRN++  +       ++ DG             TL ++S FW+ V +FLWL GEAI
Sbjct: 130 LILGFRNVIGDI-------RMFDG-----------KTLTEISQFWASVHAFLWLIGEAI 170

Query: 185 FHFLPVGIVWSVTRKMGTTQILGIVLGICLVSPQLLNAYSVASTSAADIAKNWSWNFGYF 244
           F FLPVG+ WS  +K+G T ILGI LG+ LVSPQL+NAY +             W+FG F
Sbjct: 171 FFFLPVGVCWSTVKKLGGTPILGITLGVTLVSPQLMNAYLIGKEVPE------VWDFGLF 224

Query: 245 TVQKIGYQAQVIPALLAGLSLSYLEIFWRKHIPEVVSMIFVPFLSLVPAIILAHTVLGPI 304
            ++K+GYQAQVIPA+LAG++L+++E   R+ +P  + ++ VPF+S++ +++LAH  +GP
Sbjct: 225 AIEKVGYQAQVIPAILAGVALAFIENNLRRVVPSYLYLVVVPFVSIIVSVVLAHAFIGPF 284

Query: 305 GWTLGKWISAIVLIGLTGPVKWLFGAIFGALYAPFVITGLHHMTNAIDTQLIADTKTHTT 364
           G  +G  ++      +TG    +   +FG +YAP VITG+HH TNA+D QL+ +     T
Sbjct: 285 GRVIGDGVAFAAKAAMTGDFAVIGSTLFGFMYAPLVITGIHHTTNAVDLQLMQE--LGGT 342

Query: 365 GLWPMIALSNIAQGSAVLAYYFMHRHDEKEAQISLPAAISAYLGVTEPALFGVNVKYIYP 424
            +WP+IALSNIAQ SAV+    + +  + E  IS+PAAISAYLGVTEPA++G+N+KY +P
```

```
Sbjct: 343 PIWPLIALSNIAQASAVVGIIIISK-KQGERDISVPAAISAYLGVTEPAMYGINLKYKFP 401

Query: 425 FVAGMIGSSVAGLLATTFNVQANSIGVGGLPGFLSINVKYMGYFFICMAVAIFIPLFLTL 484
           ++ MIGS++A  +  +  V AN IGVGGLPG LSI  ++   + + M +AI +P  LTL
Sbjct: 402 MLSAMIGSALAAAVCGSAGVMANGIGVGGLPGILSIQPQFWSIYLVAMLIAILVPAALTL 461

Query: 485 FFKK 488
             K
Sbjct: 462 LMYK 465
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 485> which encodes the amino acid sequence <SEQ ID 486>. Analysis of this protein sequence reveals the following:

```
     Possible site: 59
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.61   Transmembrane  466 - 482 ( 457 - 488)
   INTEGRAL    Likelihood = -8.01   Transmembrane  279 - 295 ( 275 - 306)
   INTEGRAL    Likelihood = -6.05   Transmembrane  112 - 128 ( 105 - 130)
   INTEGRAL    Likelihood = -3.35   Transmembrane  204 - 220 ( 203 - 222)
   INTEGRAL    Likelihood = -3.13   Transmembrane  255 - 271 ( 255 - 272)
   INTEGRAL    Likelihood = -2.07   Transmembrane  327 - 343 ( 325 - 344)
   INTEGRAL    Likelihood = -0.59   Transmembrane  422 - 438 ( 422 - 438)

----- Final Results -----
              bacterial membrane --- Certainty=0.4843(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF94072 GB:AE004175 PTS system, trehalose-specific IIBC
           component [Vibrio cholerae]
 Identities = 231/484 (47%), Positives = 322/484 (65%), Gaps = 28/484 (5%)

Query: 5   EQDAKSLLTAIGGKENIKVVTHCATRMRFVLNDNNKANVKEIEKISVVKGTFTNAGQFQV 64
           +QD   L+  +GG+ NI  VTHC TR+RFVLN   +A+   +E +S+VKG FTNAGQFQV
Sbjct: 10  KQDVTRLIELVGGESNIASVTHCLTRLRFVLNQPEQADKAGLEALSMVKGCFTNAGQFQV 69

Query: 65  IIGNDVPVFYNDFTAVSSIEGVSKEAAKSAAKSNQNALQRVMTMLAEIFTPIIPAIIVGG 124
           +IG +V   Y      +  + VSK+ AK AA+ N N L+R ++ LAEIF P++PAII GG
Sbjct: 70  VIGTEVDQVYKMLLEQTGKQAVSKDDAKVAARQNMNVLERGISHLAEIFVPLLPAIITGG 129

Query: 125 LILGFRNILESVPFEFLGQQVEKGKLVFDAAGDPVWNTIVRVSPFWSGVNHFLWLPGEAI 184
           LILGFRN++  +               +FD         T+  +S FW+ V+ FLWL GEAI
Sbjct: 130 LILGFRNVIGDI-------------RMFDG------KTLTEISQFWASVHAFLWLIGEAI 170

Query: 185 FHFLPVGITWSVTRKMGTTQILGIVLGICLVSPQLLNAYAVAGTPAAEIAKNWVWDFGFF 244
           F FLPVG+ WS  +K+G T ILGI LG+ LVSPQL+NAY + G    E     VWDFG F
Sbjct: 171 FFFLPVGVCWSTVKKLGGTPILGITLGVTLVSPQLMNAYLI-GKEVPE-----VWDFGLF 224

Query: 245 TINRIGYQAQVIPALLAGLSLAYLEIFWRKRIPEVVSMIFVPFLSLIPALILAHTVLGPI 304
            I ++GYQAQVIPA+LAG++LA++E   R+ +P  + ++ VPF+S+I +++LAH  +GP
Sbjct: 225 AIEKVGYQAQVIPAILAGVALAFIENNLRRVVPSYLYLVVVPFVSIIVSVVLAHAFIGPF 284

Query: 305 GWTIGKGISFVVLAGLTGPVKWLFGAIFGALYAPLVITGLHHMTNAIDTQLIADTATRTT 364
           G  IG G++F   A +TG    +   +FG +YAPLVITG+HH TNA+D QL+ +     T
Sbjct: 285 GRVIGDGVAFAAKAAMTGDFAVIGSTLFGFMYAPLVITGIHHTTNAVDLQLMQELG--GT 342

Query: 365 GLWPMIALSNIAQGSAVFAYYLMNRHEEREAEISLPAAISAYLGVTEPALFGVNVKYVYP 424
            +WP+IALSNIAQ SAV    ++++ ++ E +IS+PAAISAYLGVTEPA++G+N+KY +P
Sbjct: 343 PIWPLIALSNIAQASAVVGIIIISK-KQGERDISVPAAISAYLGVTEPAMYGINLKYKFP 401

Query: 425 FVAGMIGSGIAGLLSTTFNVQANSIGVGGLPGFMAINVKYMIPFFICMAVAIVVPMFLTF 484
           ++ MIGS +A  +  +  V AN IGVGGLPG ++I  ++   + + M +AI+VP  LT
Sbjct: 402 MLSAMIGSALAAAVCGSAGVMANGIGVGGLPGILSIQPQFWSIYLVAMLIAILVPAALTL 461

Query: 485 FFRK 488
```

```
                K
Sbjct: 462 LMYK 465
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 501/675 (74%), Positives = 573/675 (84%), Gaps = 2/675 (0%)

Query: 1   MEQFKHDAKALLEAIGGKENISAVTHCATRMRFVLNDSSKAKVKVIEELPSVKGTFTNAG 60
            M +F+ DAK+LL AIGGKENI  VTHCATRMRFVLND++KA VK IE++  VKGTFTNAG
Sbjct: 1   MGKFEQDAKSLLTAIGGKENIKVVTHCATRMRFVLNDNNKANVKEIEKISVVKGTFTNAG 60

Query: 61  QFQVIIGNDVPIFYNAFVAVSGIEGVSKEAAKSAAQKNQNPLQRVLTMLAEIFTPIIPAI 120
            QFQVIIGNDVP+FYN F AVS IEGVSKEAAKSAA+ NQN LQRV+TMLAEIFTPIIPAI
Sbjct: 61  QFQVIIGNDVPVFYNDFTAVSSIEGVSKEAAKSAAKSNQNALQRVMTMLAEIFTPIIPAI 120

Query: 121 IVGGLILGFRNILDAVPFEFLGQKVVDGVRQVDSSGHPIWNTLVDVSTFWSGVDSFLWLP 180
            IVGGLILGFRNIL++VPFEFLGQ+V  G    D++G P+WNT+V VS FWSGV+ FLWLP
Sbjct: 121 IVGGLILGFRNILESVPFEFLGQQVEKGKLVFDAAGDPVWNTIVRVSPFWSGVNHFLWLP 180

Query: 181 GEAIFHFLPVGIVWSVTRKMGTTQILGIVLGICLVSPQLLNAYSVASTSAADIAKNWSWN 240
            GEAIFHFLPVGI WSVTRKMGTTQILGIVLGICLVSPQLLNAY+VA T AA+IAKNW W+
Sbjct: 181 GEAIFHFLPVGITWSVTRKMGTTQILGIVLGICLVSPQLLNAYAVAGTPAAEIAKNWVWD 240

Query: 241 FGYFTVQKIGYQAQVIPALLAGLSLSYLEIFWRKHIPEVVSMIFVPFLSLVPAIILAHTV 300
            FG+FT+ +IGYQAQVIPALLAGLSL+YLEIFWRK IPEVVSMIFVPFLSL+PA+ILAHTV
Sbjct: 241 FGFFTINRIGYQAQVIPALLAGLSLAYLEIFWRKRIPEVVSMIFVPFLSLIPALILAHTV 300

Query: 301 LGPIGWTLGKWISAIVLIGLTGPVKWLFGAIFGALYAPFVITGLHHMTNAIDTQLIADTK 360
            LGPIGWT+GK IS +VL GLTGPVKWLFGAIFGALYAP VITGLHHMTNAIDTQLIADT
Sbjct: 301 LGPIGWTIGKGISFVVLAGLTGPVKWLFGAIFGALYAPLVITGLHHMTNAIDTQLIADTA 360

Query: 361 THTTGLWPMIALSNIAQGSAVLAYYFMHRHDEKEAQISLPAAISAYLGVTEPALFGVNVK 420
            T TTGLWPMIALSNIAQGSAV AYY M+RH+E+EA+ISLPAAISAYLGVTEPALFGVNVK
Sbjct: 361 TRTTGLWPMIALSNIAQGSAVFAYYLMNRHEEREAEISLPAAISAYLGVTEPALFGVNVK 420

Query: 421 YIYPFVAGMIGSSVAGLLATTFNVQANSIGVGGLPGFLSINVKYMGYFFICMAVAIFIPL 480
            Y+YPFVAGMIGS +AGLL+TTFNVQANSIGVGGLPGF++INVKYM  FFICMAVAI +P+
Sbjct: 421 YVYPFVAGMIGSGIAGLLSTTFNVQANSIGVGGLPGFMAINVKYMIPFFICMAVAIVVPM 480

Query: 481 FLTLFFKKSGILTKTEEEKLVPDAVIASTTETKSAKEKAVVSGTKLSVVSPLSGLAKPLD 540
            FLT FF+KS I+TKTE+E  +P+  + S      +A  K  + GT +++ SPL+G  K L
Sbjct: 481 FLTFFFRKSHIMTKTEDEAKLPETPV-SDAPVATAPHK-TMQGTVITLTSPLTGEVKALS 538

Query: 541 QASDPVFSQGIMGKGVVIDPSDGELVSPVDATVSVLFPTKHAIGLLTSEGVEFLIHIGMD 600
            +A DPVF+QG+MG+G ++ P++G LV+P DA VSVLFPTKHAI L+T+EG+E L+HIGMD
Sbjct: 539 EAVDPVFAQGVMGQGALLQPTEGVLVAPCDAEVSVLFPTKHAICLVTTEGLELLMHIGMD 598

Query: 601 TVNLEGKGFTSHVAQGDTVKVGDKLITFDIPMIKEEGYIVETPILITNQQEFRPEELIDL 660
            TVNL+G+GF + V QGD VK G  LI FDI  I E GY  ETP+++TNQ  F       L
Sbjct: 599 TVNLDGQGFEALVKQGDQVKAGQTLIQFDIAAISEAGYATETPLVVTNQDVFTVTVEGSL 658

Query: 661 PKQIKRGQALMVAKK 675
            P+QIK    L VA K
Sbjct: 659 PRQIKVNDKLAVAVK 673
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 146

A DNA sequence (GBSx0152) was identified in *S.agalactiae* <SEQ ID 487> which encodes the amino acid sequence <SEQ ID 488>. This protein is predicted to be dextran glucosidase DexS (treC). Analysis of this protein sequence reveals the following:

```
Possible site: 48
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3493(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB65079 GB:U35633 dextran glucosidase DexS [Streptococcus suis]
 Identities = 383/547 (70%), Positives = 439/547 (80%), Gaps = 13/547 (2%)

Query: 1   MTIDKRKVVYQIYPKSYKDTTGNGVGDLRGIIEKLPYLAELGIDMVWLNPFYPSPQRDNG 60
           MTIDKRKVVYQIYPKSYKDTTGNGVGDLRGIIEKLPYL ELGIDM+WLNPFYPSPQRDNG
Sbjct: 1   MTIDKRKVVYQIYPKSYKDTTGNGVGDLRGIIEKLPYLKELGIDMIWLNPFYPSPQRDNG 60

Query: 61  YDISDYTAINPDFGTMDDFEEMIEVGRQYRIDFMLDMVLNHCSIEHEWFKKALAGDRYYQ 120
           YDISDYTA+NPDFGTM DFEEM+ VG++  I+FMLDMVLNHCS +HEWF+KAL+GD+YYQ
Sbjct: 61  YDISDYTAVNPDFGTMADFEEMVTVGKELGIEFMLDMVLNHCSTDHEWFQKALSGDQYYQ 120

Query: 121 DFFILRDNPTDWVSKFGGNAWAPFGDTGKYYLHLFDITQADLNWRNADVRKELFKVVNFW 180
           DFFILRD PTDWVSKFGGNAWAPFGDTGKYYLHLFD+TQADLNWRN  +R+ELFKVVNFW
Sbjct: 121 DFFILRDQPTDWVSKFGGNAWAPFGDTGKYYLHLFDVTQADLNWRNPHIREELFKVVNFW 180

Query: 181 RDKGVKGFRFDVINLIGKDEILENCPINDGKPAYTDRPITHDYLKMLNNASFGQDDSFMT 240
           +DKGVKGFRFDVINLIGKDE  E+CPINDGKPAYTDRPITHDYLKM+NNA+FG +  FMT
Sbjct: 181 KDKGVKGFRFDVINLIGKDEAREDCPINDGKPAYTDRPITHDYLKMMNNATFGSEKGFMT 240

Query: 241 VGEMSSTTIANCILYTAPEREELSMAFNFHHLKVDYKDGQKWTIMAFDFPALRDLFHSWG 300
           VGEMS+TTI NCILYTAPER+ELSMAFNFHHLKVDYKDGQKWTIM FDF  L+ LFH+WG
Sbjct: 241 VGEMSATTIENCILYTAPERKELSMAFNFHHLKVDYKDGQKWTIMDFDFEELKHLFHTWG 300

Query: 301 EGMSEGNGWNALFYNNHDQPRALNRFVDVKRFRNEGATMLAASIHLSRGTPYIYMGEEIG 360
           E MS GNGWNALFYNNHDQPRALNRF+DV+ FR EGATMLAASIHLSRG
Sbjct: 301 EEMSVGNGWNALFYNNHDQPRALNRFIDVENFRKEGATMLAASIHLSRGNNLTST----- 355

Query: 361 MLDPDYSSMDDYVDIESLNAYQIMLDEGKSQEEAFSIIRAKSRDNSRVPMQWDDS----- 415
            +    SS    +   + + +    S  +   + R  SR +   P+
Sbjct: 356 WVRRSVSSTLTTIAWTTTWTWSLSMPTRCSWTKVTRLSR-PSRLSRPSPVTIPAPRCNGT 414

Query: 416 --TNAGFSEGAPWLKVGKSYKEINVAKEKTGLIFTFYQELIRLRKQLPIIADGNYKAAFK 473
             T     +  PWLK GKSY+ INV +EKTG IFTFY+    LRK+LP+I++G+YKAA+K
Sbjct: 415 LLTMQASQQATPWLKAGKSYQTINVEQEKTGPIFTFYKRTHPLRKELPLISEGDYKAAYK 474

Query: 474 DNEKVYAFERHLDKEKLLVLNNFFAEKVKIKLPENYLQGQVLLSNYKDVTLDETVTLQPY 533
           D++KVYAFER L+ EKLLVLNNFFAE+V++ L ++Y  GQVL+SNY D  L + + L+PY
Sbjct: 475 DSQKVYAFERLLNDEKLLVLNNFFAEEVELDLADDYAHGQVLISNYPDNKLGKKIILKPY 534

Query: 534 QTLAILV 540
           Q LAI V
Sbjct: 535 QALAIQV 541
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 489> which encodes the amino acid sequence <SEQ ID 490>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3631(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 431/539 (79%), Positives = 486/539 (89%)
```

```
Query: 1   MTIDKRKVVYQIYPKSYKDTTGNGVGDLRGIIEKLPYLAELGIDMVWLNPFYPSPQRDNG 60
           MTIDK+KVVYQIYPKSYKDTTGNGVGDL GII+KLPYL ELGIDM+WLNPFYPSPQRDNG
Sbjct: 1   MTIDKKKVVYQIYPKSYKDTTGNGVGDLLGIIDKLPYLQELGIDMIWLNPFYPSPQRDNG 60

Query: 61  YDISDYTAINPDFGTMDDFEEMIEVGRQYRIDFMLDMVLNHCSIEHEWFKKALAGDRYYQ 120
           YD+SDYTA+NPDFGTM DFE +++  ++++I+ MLDMVLNHCS +HEWF+KALAGD YYQ
Sbjct: 61  YDVSDYTAVNPDFGTMADFENLVKAAKEHQIELMLDMVLNHCSTDHEWFQKALAGDPYYQ 120

Query: 121 DFFILRDNPTDWVSKFGGNAWAPFGDTGKYYLHLFDITQADLNWRNADVRKELFKVVNFW 180
           DFFILRD PTDWVSKFGGNAWAPFGDTGKYYLHLFD+TQADLNWRN  VR+EL KVVNFW
Sbjct: 121 DFFILRDQPTDWVSKFGGNAWAPFGDTGKYYLHLFDVTQADLNWRNPHVREELAKVVNFW 180

Query: 181 RDKGVKGFRFDVINLIGKDEILENCPINDGKPAYTDRPITHDYLKMLNNASFGQDDSFMT 240
           RDKGVKGFRFDVINLIGKDE L +CP+NDGKPAYTDRPITH YL  LN ASFGQDDSFMT
Sbjct: 181 RDKGVKGFRFDVINLIGKDEELVDCPVNDGKPAYTDRPITHTYLHDLNQASFGQDDSFMT 240

Query: 241 VGEMSSTTIANCILYTAPEREELSMAFNFHHLKVDYKDGQKWTIMAFDFPALRDLFHSWG 300
           VGEMS+TTI NC+LYTAPEREELSMAFNFHHLKVDY++GQKWTIMAFDF ALRDLFH+WG
Sbjct: 241 VGEMSATTIDNCLLYTAPEREELSMAFNFHHLKVDYENGQKWTIMAFDFAALRDLFHAWG 300

Query: 301 EGMSEGNGWNALFYNNHDQPRALNRFVDVKRFRNEGATMLAASIHLSRGTPYIYMGEEIG 360
           EGMS+GNGWNALFYNNHDQPRALNRFVDV  FRNEGATMLAASIHLSRGTPYIYMGEEIG
Sbjct: 301 EGMSQGNGWNALFYNNHDQPRALNRFVDVTHFRNEGATMLAASIHLSRGTPYIYMGEEIG 360

Query: 361 MLDPDYSSMDDYVDIESLNAYQIMLDEGKSQEEAFSIIRAKSRDNSRVPMQWDDSTNAGF 420
           MLDPD+ SMDDYVD+ESLNAY  +L  GKS EEAF+II+AKSRDN+R PMQWD S +AGF
Sbjct: 361 MLDPDFDSMDDYVDVESLNAYSSLLVSGKSAEEAFAIIKAKSRDNARTPMQWDASEHAGF 420

Query: 421 SEGAPWLKVGKSYKEINVAKEKTGLIFTFYQELIRLRKQLPIIADGNYKAAFKDNEKVYA 480
           + G PWL+VGKSY++INV  EK G IF FYQ LI LRK+LPIIA+G+Y+AAFKD++ VYA
Sbjct: 421 TTGKPWLEVGKSYRDINVETEKEGRIFPFYQRLIALRKELPIIAEGDYRAAFKDSQAVYA 480

Query: 481 FERHLDKEKLLVLNNFFAEKVKIKLPENYLQGQVLLSNYKDVTLDETVTLQPYQTLAIL 539
           FERHL  + LLVLN+F+A++V+++LP  Y  GQVL+SNY+ V++ E V L+PYQTLAIL
Sbjct: 481 FERHLGDQCLLVLNHFYADEVELELPPRYQHGQVLISNYEKVSICEKVILKPYQTLAIL 539
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 147

A DNA sequence (GBSx0153) was identified in *S.agalactiae* <SEQ ID 491> which encodes the amino acid sequence <SEQ ID 492>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -3.03   Transmembrane    8 -  24 (   8 -  25)

----- Final Results -----
               bacterial membrane --- Certainty=0.2211(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 148

A DNA sequence (GBSx0154) was identified in *S.agalactiae* <SEQ ID 493> which encodes the amino acid sequence <SEQ ID 494>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB03939 GB:AP001507 unknown conserved protein [Bacillus halodurans]
 Identities = 190/639 (29%), Positives = 331/639 (51%), Gaps = 34/639 (5%)

Query: 6   TVVIMLVFLARKNLSLYELTVQTKFSIKVIIEQINYLNSFLAKNHLPAIAHSAGRYQLLG 65
            T ++  +  AR  L + ELT +   S + +   +  +NS+L +  L A+ +      L+
Sbjct: 8   TFILTQLLHARSYLPIQELTQKLNVSRRTVYNDLEKINSWLEEQGLKAV-YKVRSQGLIL 66

Query: 66  DEKEHDKI---VSLLEAEQFYLTQEERVCLIYLYSFCRREFVSNVHYQDFLKVSKNTTLS 122
            DE+  ++I    +  L++  +  + +ER   + +Y   R E +    H  D   VS+NTT+
Sbjct: 67  DERAKEEIPTKLRSLKSWHYEYSAQERKAWVVIYLLTRLEPLFLEHLMDRTGVSRNTTID 126

Query: 123 DIKMLRSKLAKRGISLTYTRAKGYSLVGDEMDKHQVAFQMITQLLE--------SPIGFW 174
            DIK L+ +L     ++L + R  GY++ GDE DK +      ++Q L          SPI  +
Sbjct: 127 DIKCLKDELNNFHLALEFERKDGYTISGDETDKRKALVYYLSQALPQQNWETELSPIRIF 186

Query: 175 SLNYILSSWKFALSYEKLEKTVEYFYESFQLSPIQ---DRLEKSLYFIILILCRYQRSVD 231
                  +    F +  E+L+K  +   ES ++  IQ   D L        +L + R  +
Sbjct: 187 LRTKRDNGRIFTI--EELQKVYDVISESEKVLKIQYTDDVLHSLSLRFLLFMKRVAKG-- 242

Query: 232 RVLQGSPIVSEQLK-----ELTTIIVTNLSQDISLSKPLDQKEKDYITLILSGCF----- 281
            + ++  P+  + LK      E   ++   L Q   +  P D++     T ILS
Sbjct: 243 KFIKVHPLEKQVLKGTKEYEAAKVMSFKLEQAFGVHYP-DEEVLYLTTHILSSKINYANG 301

Query: 282 EGEGTKDDDFFEALAKAIVDEMETVSLLNFSNKEELLQGLKRHIIPAYFRLKYGLTGDSG 341
            E E  K+        +  ++V++ +  + + F  KE L + L  HI  PA++R+KYGL  ++
Sbjct: 302 EIESRKESQELTHIVTSMVNDFQKYACVVFEEKELLEKNLFFHIKPAFYRIKYGLEVENN 361

Query: 342 YTQNIKEHYSDLFLLVKKALRPLEEQVGL-IPDSEISYFVIHFGGYLRQSGGTQSMSYKA 400
              ++IK   Y +LFLL +K +   LE  VG  + D+E+++  +HF  G++R+ G    +   KA
Sbjct: 362 IAESIKTSYPELFLLTRKVVHYLERYVGKSVNDNEVAFITMHFVGWMRREGTIPTKRKKA 421

Query: 401 LILCPNGVSSSLVIKEKLRGLFPQIHFHRVSKIEQLKLIDNQTYDMVFSTIFVETKKPNY 460
            LI+C NGV +S   +K +L GLFP +   +   I + +    +  ++ +T    E   P +
Sbjct: 422 LIVCANGVGTSQFLKNQLEGLFPAVDIIKTCSIREYEKTPVEVDFIISTTSIPEKNVPIF 481

Query: 461 LVSLMMT-AEQVQQLKELVISDFPKACLDDFQLDQLIATIKKYAHVHCEEELKLALRTMV 519
            +V+  ++T   E+ + LK + ++       +  + ++ L+  IK++ +V  E+ L    LR
Sbjct: 482 IVNPILTETEKERLLKSVHVALDELGAMKGYSIEGLMDVIKRHGNVDDEKALYQDLRRFF 541

Query: 520 KQD--ILRKDVRPLLHQLITEETYQTSSEQMNWKEAIRLAAKPLLASGKITESYPEAMIE 577
              Q    I  K  +P L+QL+TE+  Q    +  +W+EAI+LAAKPLL  G +TESY + MI+
Sbjct: 542 TQPTPIGPKQEKPDLNQLLTEDMIQLREQVTHWQEAIQLAAKPLLLKGMVTESYVKKMIK 601

Query: 578 KVEEFGPFINLGKGIAIPHARPEDGVNSVGMSMLVLEQP 616
             +E+FGP++ +    AIPHA+PEDGV  +GMS+L L++P
Sbjct: 602 NIEKFGPYMIIAPHFAIPHAKPEDGVRQLGMSLLWLKKP 640
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 495> which encodes the amino acid sequence <SEQ ID 496>. Analysis of this protein sequence reveals the following:

```
Possible site: 57 or 61
>>> Seems to have no N-terminal signal sequence
```

```
   INTEGRAL    Likelihood = -0.64   Transmembrane  123 - 139 ( 123 - 139)

----- Final Results -----
               bacterial membrane --- Certainty=0.1256(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 187/624 (29%), Positives = 327/624 (51%), Gaps = 20/624 (3%)

Query: 1   MVDNKTVVIMLVFLARKNLSLYELTVQTKFSIKVIIEQINYLNSFLAKNHLPAIAHSAGR 60
           M+ ++ +    +F   K  SL       K S + I+  I  +N  L+   LP IA
Sbjct: 35  MLSHELIRNYQLFSKYKGHSLEAFESILKASKRHILADIAKINDTLSLYQLPLIALDR-- 92

Query: 61  YQLL--GDEKEHDKIVSLLEAEQFYLTQEERVCLIYLYSFCRREFVSNVHYQDFLKVSKN 118
            QL+   D  E D +  +L     YL Q+ER+ +I +Y    +EF+S  H +  L++S+N
Sbjct: 93  -QLVYPPDLTEKDLLNRMLPTLDDYLFQDERLDMIIIYIMMAKEFISINHLESLLRLSRN 151

Query: 119 TTLSDIKMLRSKLAKRGISLTYTRAKGYSLVGDEMDKHQVAFQMITQLLESPIGFWSLNY 178
           + ++D+ ++R ++    ++L Y R  GY   G+ +   ++    ++ LL+   G W  +Y
Sbjct: 152 SVIADLNLVRDRVQAFQVTLAYNRQDGYFFEGEPLALRRLLESAVSSLLQVTSGPWVFSY 211

Query: 179 ILSSWKFALSYEKLEKTVEYFYESFQLSPIQDRLEKSLYFIILILCR-YQRSVD-RVLQG 236
           +L         + +  T+E       L+ I ++L   +YF  L+  R + R+V    +
Sbjct: 212 LLHELGLPDQKKVMAATLEELSRENHLTFISEKLRDLIYFFCLLAHRPFSRNVRAEAVDT 271

Query: 237 SPIVSEQLKELTTIIVTNLSQDISLSKPLDQKEKDYITLILSGCFEG--EGTKDDDFFEA 294
            P+ S  ++ +   ++ N          P   +EK  +   L GC +G  E        ++
Sbjct: 272 FPLASPAVETMVDQLLVNF--------PSLTEEKYLVQSRLLGCIQGDLELVFQQPIYDI 323

Query: 295 LAKAIVDEMETVSLLNFSNKEELLQGLKRHIIPAYFRLKYGLTGDSGYTQNIKEHYSDLF 354
           + + I++ +   + L+ ++  EL Q L  H++PAY+RL Y +    +   + IK+ Y  LF
Sbjct: 324 MEE-IINSVAVNTGLSITDTPELRQNLYSHLLPAYYRLYYDINLTNPLKEQIKQDYESLF 382

Query: 355 LLVKKALRPLEEQVGL-IPDSEISYFVIHFGGYLRQSGGTQSMSYKALILCPNGVSSSLV 413
            LVK++L PLE+Q+G  + + E++YF IHFG +L+       S    AL +CPNG+SSSL+
Sbjct: 383 YLVKRSLSPLEKQLGKSVNEDEVAYFTIHFGRWLQAPKKRPSNQLVALSVCPNGISSSLM 442

Query: 414 IKEKLRGLFPQIHFHRVSKIEQLKLIDNQTYDMVFSTIFVETKKPNYLVSLMMTAEQVQQ 473
           ++  L+ LFPQ+ F R+ +++++KL+D  ++D++FST+  + KP Y+   +M   +
Sbjct: 443 LEATLKELFPQLQFIRIHQLDKIKLLDPASFDLIFSTVAFDCAKPVYVTQALMGPVEKMM 502

Query: 474 LKELVISDFPKACLDDFQLDQLIATIKKYAHVHCEEELKLAL-RTMVKQDILRKDVRPLL 532
           LK++V  DF     + F LD L++ I K+  +  +E L   L R ++   +  +     L
Sbjct: 503 LKKMVCDDFHLPLSEQFALDDLLSIIHKHTTITNKEGLVSDLSRYLIGNHLTIEKGGLGL 562

Query: 533 HQLITEETYQTSSEQMNWKEAIRLAAKPLLASGKITESYPEAMIEKVEEFGPFINLGKGI 592
           H L+T +  + +E  +W+EAIRLAA+PLL    I  SY + MI+ V E G +I L    +
Sbjct: 563 HDLLTADFIRQADAVSDWQEAIRLAAQPLLEHQMIETSYIDGMIDSVNELGAYIVLAPKV 622

Query: 593 AIPHARPEDGVNSVGMSMLVLEQP 616
           A+PHA PE G   +GMS+L L++P
Sbjct: 623 AVPHAAPEKGTRQLGMSLLQLKEP 646
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 149

A DNA sequence (GBSx0155) was identified in *S.agalactiae* <SEQ ID 497> which encodes the amino acid sequence <SEQ ID 498>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3665(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 499> which encodes the amino acid sequence <SEQ ID 500>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3665(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 33/35 (94%), Positives = 35/35 (99%)

Query: 1  MEKEAKQIIDLKRNLFKIDVRAQKDEEKVFMRTAW 35
          +EKEAKQ+IDLKRNLFKIDVRAQKDEEKVFMRTAW
Sbjct: 1  LEKEAKQMIDLKRNLFKIDVRAQKDEEKVFMRTAW 35
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 150

A repeated DNA sequence (GBSx0156) was identified in *S.agalactiae* <SEQ ID 501> which encodes the amino acid sequence <SEQ ID 502>. This protein is predicted to be a repeat-associated protein in rhsc-phrb intergenic region. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.57   Transmembrane   29 -  45 (  28 -  48)

----- Final Results -----
               bacterial membrane --- Certainty=0.2826(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A closely-related DNA sequence was identified in *S.agalactiae* <SEQ ID 1035> which encodes the amino acid sequence <SEQ ID 1036>. Further related GBS sequences are: <SEQ ID 9067>, <SEQ ID 9068>, <SEQ ID 9497>, <SEQ ID 9498>, <SEQ ID 9733>, <SEQ ID 9734>

A related repeated DNA sequence was identified in *S.pyogenes* <SEQ ID 503> which encodes the amino acid sequence <SEQ ID 504>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.57   Transmembrane   29 -  45 (  28 -  48)

----- Final Results -----
               bacterial membrane --- Certainty=0.2826(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS gene <SEQ ID 8547> and protein <SEQ ID 8548> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 5
McG: Discrim Score:     -7.73
GvH: Signal Score (-7.5): -3.88
     Possible site: 44
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -4.57 threshold:  0.0
   INTEGRAL    Likelihood = -4.57   Transmembrane   26 -  42 (  25 -  45)
   PERIPHERAL  Likelihood =  2.12     334
 modified ALOM score:   1.41

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.2826(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7071> which encodes the amino acid sequence <SEQ ID 7072>. An alignment of the GAS and GBS sequences follows:

```
 Score =  767 bits (1960), Expect = 0.0
 Identities = 375/377 (99%), Positives = 375/377 (99%)

Query: 4   MIDFIISIDDCAVELDSRQSWKIRSPLSTILFLVFVCQLAGIETWKEMEDFIEMNEPLFA 63
           MIDFIISIDDCAVELDSRQSWKIR PLSTILFLVFVCQLAGIETWKEMEDFIEMNEPLFA
Sbjct: 1   MIDFIISIDDCAVELDSRQSWKIRYPLSTILFLVFVCQLAGIETWKEMEDFIEMNEPLFA 60

Query: 64  TYVDLSEGCSSHDTLERVISLVNSDRLKELKVQFEQSLTSLDAVHQLISVDGKTIRGNRG 123
           TYVDLSEGC SHDTLERVISLVNSDRLKELKVQFEQSLTSLDAVHQLISVDGKTIRGNRG
Sbjct: 61  TYVDLSEGCPSHDTLERVISLVNSDRLKELKVQFEQSLTSLDAVHQLISVDGKTIRGNRG 120

Query: 124 KNQKPVHIVTAYDGGHHLSLGQVAVEEKSNEIVAIPQLLRTIDIRKSIVTIDAMGTQTAI 183
           KNQKPVHIVTAYDGGHHLSLGQVAVEEKSNEIVAIPQLLRTIDIRKSIVTIDAMGTQTAI
Sbjct: 121 KNQKPVHIVTAYDGGHHLSLGQVAVEEKSNEIVAIPQLLRTIDIRKSIVTIDAMGTQTAI 180

Query: 184 VDTIIKGKADYCLAVKGNQETLYDDIALYFSDVNLLEELQENAQYYQTVEKSRGQIEVRE 243
           VDTIIKGKADYCLAVKGNQETLYDDIALYFSDVNLLEELQENAQYYQTVEKSRGQIEVRE
Sbjct: 181 VDTIIKGKADYCLAVKGNQETLYDDIALYFSDVNLLEELQENAQYYQTVEKSRGQIEVRE 240

Query: 244 YWVSSDIKWLCQNHPKWHKLRGIGMTRNTIDKDGQLSQENRYFIFSFKPDVLTFANCVRG 303
           YWVSSDIKWLCQNHPKWHKLRGIGMTRNTIDKDGQLSQENRYFIFSFKPDVLTFANCVRG
Sbjct: 241 YWVSSDIKWLCQNHPKWHKLRGIGMTRNTIDKDGQLSQENRYFIFSFKPDVLTFANCVRG 300

Query: 304 HWQIESMHWLLDVVYHEDHHQTLDKRAAFNLNLIRKMCLYFLKVMVFPKKDLSYRRKQRY 363
           HWQIESMHWLLDVVYHEDHHQTLDKRAAFNLNLIRKMCLYFLKVMVFPKKDLSYRRKQRY
Sbjct: 301 HWQIESMHWLLDVVYHEDHHQTLDKRAAFNLNLIRKMCLYFLKVMVFPKKDLSYRRKQRY 360

Query: 364 ISVHLEDYLVQLFGERG 380
           ISVHLEDYLVQLFGERG
Sbjct: 361 ISVHLEDYLVQLFGERG 377
```

A further related DNA sequence was identified in *S.pyogenes* <SEQ ID 9087> which encodes the amino acid sequence <SEQ ID 9088>. A further related DNA sequence was identified in *S.pyogenes* <SEQ ID 9089> which encodes the amino acid sequence <SEQ ID 9090>. The GAS and GBS proteins are 100% identical.

There is also homology to SEQ IDs 7018 and 8548.

SEQ ID 8548 (GBS318) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 46 (lane 5; MW 70kDa).

GBS318-GST was purified as shown in Figure 203, lane 3.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 151

A DNA sequence (GBSx0157) was identified in *S.agalactiae* <SEQ ID 505> which encodes the amino acid sequence <SEQ ID 506>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 496.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 152

A repeated DNA sequence (GBSx0158) was identified in *S.agalactiae* <SEQ ID 507> which encodes the amino acid sequence <SEQ ID 508>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1054(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB03941 GB:AP001507 unknown conserved protein [Bacillus halodurans]
 Identities = 26/82 (31%), Positives = 52/82 (62%), Gaps = 2/82 (2%)

Query: 2  LRIGTACGSGLGSSFMVQMNIESILKDLGVSDVEVEHYDLGGADPSAADVWIVGRDLEDS 61
          ++I   CG G G+S +++MN+E++L  LG++  +V++ D+  A    +D  I  ++L +S
Sbjct: 1  MKILCVCGLGQGTSLILKMNVETVLSQLGIA-ADVDNTDVSSASSEQSDFIITSKELAES 59

Query: 62 -AGHLGDVRILNSIIDMDELRE 82
            A H   + I+N+  DM+E+++
Sbjct: 60 LASHPSKIVIVNNYFDMEEIKQ 81
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 509> which encodes the amino acid sequence <SEQ ID 510>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 27/90 (30%), Positives = 51/90 (56%), Gaps = 1/90 (1%)

Query: 1  MLRIGTACGSGLGSSFMVQMNIESILKDLGVSDVEVEHYDLGGADPSAADVWIVGRDLED 60
          M++I T CG+G+GSS +++M +E+I   LG+ DV+ E  D   A    AD+++  ++ +D
Sbjct: 8  MIKIVTVCGNGIGSSLLLRMKVEAIASSLGI-DVDAESCDSNAAVGKGADLFVTVKEFKD 66

Query: 61 SAGHLGDVRILNSIIDMDELRELVTGICQE 90
                V I+ S  +  ++ E +  + +E
Sbjct: 67 IFPEDAKVCIVKSYTNRKKIEEDLVPVLKE 96
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 153

A DNA sequence (GBSx0159) was identified in *S.agalactiae* <SEQ ID 511> which encodes the amino acid sequence <SEQ ID 512>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 154

A DNA sequence (GBSx0160) was identified in *S.agalactiae* <SEQ ID 513> which encodes the amino acid sequence <SEQ ID 514>. This protein is predicted to be sgaT. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-14.97   Transmembrane  424 - 440 ( 411 - 447)
   INTEGRAL    Likelihood = -8.86   Transmembrane  224 - 240 ( 221 - 248)
   INTEGRAL    Likelihood = -7.27   Transmembrane  134 - 150 ( 124 - 167)
   INTEGRAL    Likelihood = -7.11   Transmembrane  321 - 337 ( 314 - 349)
   INTEGRAL    Likelihood = -6.64   Transmembrane  379 - 395 ( 370 - 397)
   INTEGRAL    Likelihood = -6.21   Transmembrane   96 - 112 (  94 - 115)
   INTEGRAL    Likelihood = -6.05   Transmembrane  267 - 283 ( 257 - 289)
   INTEGRAL    Likelihood = -3.13   Transmembrane   18 -  34 (  17 -  35)
   INTEGRAL    Likelihood = -2.55   Transmembrane  151 - 167 ( 151 - 167)
   INTEGRAL    Likelihood = -0.32   Transmembrane   42 -  58 (  42 -  58)

----- Final Results -----
               bacterial membrane --- Certainty=0.6986(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB52363 GB:AL109747 putative integral membrane protein
          [Streptomyces coelicolor A3(2)]
 Identities = 202/453 (44%), Positives = 292/453 (63%), Gaps = 22/453 (4%)
```

```
Query: 7   FLVN-IASTPAILVALIAIIGLVLQKKGVPDIVKGGIKTFVGFLVVSGGTGIVQNSLNPF 65
           FLVN I S PA L+ +I  +GL   KK V   V G IK  +G L+V  G G+V +SL+P
Sbjct: 10  FLVNEILSQPAYLIGIITAVGLAALKKSVGQTVGGAIKATLGLLLVGAGAGLVSSSLDPL 69

Query: 66  GKMFEHAFHLVGVVPNNEAIVAVALTKYGSATALIMLAGMIFNILIARFTKFKYIFLTGH 125
           G+M +      GV+P NEAIV +A +++G+  A +M+ G + ++ +ARFT  +Y+FLTGH
Sbjct: 70  GRMIQGTTGTHGVIPTNEAIVGIAQSEFGARVAWLMILGFLVSLALARFTPLRYVFLTGH 129

Query: 126 HTLYMACMIAVIFAVAGFTSFSLILFGGLALGIIMSVSPAFVQKYMIQLTGNDKVALGHF 185
           H L+MA ++ ++ A AG  S +++L GG+ +GI++   PAF   +  ++TGND +A+GHF
Sbjct: 130 HMLFMATLLTIVMATAGQGSVAVVLGGGVLVGILLVALPAFAHPWTKKVTGNDTLAIGHF 189

Query: 186 GSLGYWLSGFIGGIVGDKSKSTEDIKFPKSLSFLRDSTVSITISMAIIYLIVAV------ 239
           G+ GY +SG  G +VG  S+STE++K P+ L FLRDS V+  +SM +IYL++++
Sbjct: 190 GTAGYIVSGATGQLVGKNSRSTEEMKLPEGLRFLRDSMVATALSMVLIYLVMSLLFLAKV 249

Query: 240 --------FAGEAYIAKEISNGVNGLVYALQLAGQFAAGVFVILAGVRLILGEIVPAFKG 291
                FAG        ++  N L+ ++    QF  GV VIL GVR ILGE+VPAF+G
Sbjct: 250 GQDAAFKAFAGSG--GDPAADVGNYLMQSVMQGLQFGIGVAVILFGVRTILGELVPAFQG 307

Query: 292 ISEKLVPNSKPALDCPIVYPYAPNAVLIGFISSFVGGLVSMIVMI-----VTGTTVILPG 346
           I+ ++VP +KPALD PIV+PYA NAVLIGFI SF+GGL  +  +I       G  ++LPG
Sbjct: 308 IAGRVVPGAKPALDAPIVFPYAQNAVLIGFIFSFLGGLTGLAALIWVFNPAFGLALVLPG 367

Query: 347 VVPHFFCGATAGVIGNASGGVRGATIGAFVQGILISFLPIFLMPVLGGLGFKGSTFSDAD 406
           +VPHFF G  AGV GNA+GG RGA +G+F+ G+LI+FLP  L+  LG  G   +TF DAD
Sbjct: 368 LVPHFFTGGAAGVYGNATGGRRGAAVGSFLNGLLITFLPAILLKALGSFGEANTTFGDAD 427

Query: 407 FGLTGIILGALNHVGGAIAIVIGIVVILIGLFG 439
           FG  G +LG++  +G   ++  ++  L+ L G
Sbjct: 428 FGWFGAVLGSIGKLDGTAGLIGMLIFGLLILAG 460
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 515> which encodes the amino acid sequence <SEQ ID 516>. Analysis of this protein sequence reveals the following:

```
     Possible site: 34
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -8.33   Transmembrane  330 - 346 ( 315 - 353)
   INTEGRAL    Likelihood = -8.17   Transmembrane  227 - 243 ( 221 - 246)
   INTEGRAL    Likelihood = -4.62   Transmembrane  127 - 143 ( 126 - 145)
   INTEGRAL    Likelihood = -4.25   Transmembrane  269 - 285 ( 266 - 291)
   INTEGRAL    Likelihood = -3.77   Transmembrane   43 -  59 (  41 -  62)
   INTEGRAL    Likelihood = -3.66   Transmembrane   98 - 114 (  91 - 116)
   INTEGRAL    Likelihood = -2.76   Transmembrane  146 - 162 ( 145 - 163)
   INTEGRAL    Likelihood = -1.59   Transmembrane  308 - 324 ( 308 - 324)

----- Final Results -----
              bacterial membrane --- Certainty=0.4333(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB52363 GB:AL109747 putative integral membrane protein
          [Streptomyces coelicolor A3(2)]
 Identities = 162/387 (41%), Positives = 245/387 (62%), Gaps = 17/387 (4%)

Query: 8   IRDILKEPAFLMGLIAFAGLVALKTPAHKVLTGTLGPILGYLMLVAGAGVIVTNLDPLAK 67
           + +IL +PA+L+G+I   GL ALK    + + G +   LG L++ AGAG++ ++LDPL +
Sbjct: 12  VNEILSQPAYLIGIITAVGLAALKKSVGQTVGGAIKATLGLLLVGAGAGLVSSSLDPLGR 71

Query: 68  LIEHGFSITGVVPNNEAVTSVAQKILGVETMSILVVGLLLNLAFARFTRFKYIFLTGHHS 127
           +I+      GV+P NEA+  +AQ   G     ++++G L++LA ARFT  +Y+FLTGHH
Sbjct: 72  MIQGTTGTHGVIPTNEAIVGIAQSEFGARVAWLMILGFLVSLALARFTPLRYVFLTGHHM 131

Query: 128 FFMACLLSAVLGAVGFKGSLLIIL-DGFLLGAWSAISPAIGQQYTLKVTDGDEIAMGHFG 186
            FMA LL+ V+   G +GS+ ++L  G L+G      PA    +T KVT  D +A+GHFG
```

```
Sbjct: 132 LFMATLLTIVMATAG-QGSVAVVLGGGVLVGILLVALPAFAHPWTKKVTGNDTLAIGHFG 190

Query: 187 SLGYYLSAWVGSKVGKDSKDTEDLQISEKWSFLRNTTISTGLIMVIFYLVAT---VASVL 243
           + GY +S   G  VGK+S+ TE++++ E   FLR++ ++T L MV+ YLV +   +A V
Sbjct: 191 TAGYIVSGATGQLVGKNSRSTEEMKLPEGLRFLRDSMVATALSMVLIYLVMSLLFLAKVG 250

Query: 244 RNASVAEELAAGQNP-------FIFAIKSGLTFAVGVAIVYAGVRMILADLIPAFQGIAN 296
           ++A+      +G +P        + ++  GL F +GVA++  GVR IL +L+PAFQGIA
Sbjct: 251 QDAAFKAFAGSGGDPAADVGNYLMQSVMQGLQFGIGVAVILFGVRTILGELVPAFQGIAG 310

Query: 297 KLIPNAIPAVDCAVFFPYAPTAVIIGFASSFVGGLLGMLIL-----GVAGGVLIIPGMVP 351
           +++P A PA+D  + FPYA  AV+IGF  SF+GGL G+  L        G  L++PG+VP
Sbjct: 311 RVVPGAKPALDAPIVFPYAQNAVLIGFIFSFLGGLTGLAALIWVFNPAFGLALVLPGLVP 370

Query: 352 HFFCGATAEIFGNSTGGRRGAMIGASL 378
           HFF G  A ++GN+TGGRRGA +G+ L
Sbjct: 371 HFFTGGAAGVYGNATGGRRGAAVGSFL 397
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 174/376 (46%), Positives = 258/376 (68%), Gaps = 2/376 (0%)

Query: 1   MKGLLDFLVNIASTPAILVALIAIIGLVLQKKGVPDIVKGGIKTFVGFLVVSGGTGIVQN 60
           M+ LL F+ +I   PA L+ LIA  GLV  K     ++ G +   +G+L++  G G++
Sbjct: 1   MEALLSFIRDILKEPAFLMGLIAFAGLVALKTPAHKVLTGTLGPILGYLMLVAGAGVIVT 60

Query: 61  SLNPFGKMFEHAFHLVGVVPNNEAIVAVALTKYGSATALIMLAGMIFNILIARFTKFKYI 120
           +L+P  K+ EH F + GVVPNNEA+ +VA    G  T  I++ G++ N+  ARFT+FKYI
Sbjct: 61  NLDPLAKLIEHGFSITGVVPNNEAVTSVAQKILGVETMSILVVGLLLNLAFARFTRFKYI 120

Query: 121 FLTGHHTLYMACMIAVIFAVAGFTSFSLILFGGLALGIIMSVSPAFVQKYMIQLTGNDKV 180
           FLTGHH+ +MAC+++ +    GF    LI+  G  LG   ++SPA  Q+Y +++T  D++
Sbjct: 121 FLTGHHSFFMACLLSAVLGAVGFKGSLLIILDGFLLGAWSAISPAIGQQYTLKVTDGDEI 180

Query: 181 ALGHFGSLGYWLSGFIGGIVGDKSKSTEDIKFPKSLSFLRDSTVSITISMAIIYLI--VA 238
           A+GHFGSLGY+LS ++G  VG  SK TED++  + SFLR++T+S  + M I YL+  VA
Sbjct: 181 AMGHFGSLGYYLSAWVGSKVGKDSKDTEDLQISEKWSFLRNTTISTGLIMVIFYLVATVA 240

Query: 239 VFAGEAYIAKEISNGVNGLVYALQLAGQFAAGVFVILAGVRLILGEIVPAFKGISEKLVP 298
                A +A+E++ G N  ++A++    FA GV ++ AGVR+IL  +++PAF+GI+ K L+P
Sbjct: 241 SVLRNASVAEELAAGQNPFIFAIKSGLTFAVGVAIVYAGVRMILADLIPAFQGIANKLIP 300

Query: 299 NSKPALDCPIVYPYAPNAVLIGFISSFVGGLVSMIVMIVTGTTVILPGVVPHFFCGATAG 358
           N+ PA+DC + +PYAP AV+IGF SSFVGGL+ M+++ V G  +I+PG+VPHFFCGATA
Sbjct: 301 NAIPAVDCAVFFPYAPTAVIIGFASSFVGGLLGMLILGVAGGVLIIPGMVPHFFCGATAE 360

Query: 359 VIGNASGGVRGATIGA 374
           + GN++GG RGA IGA
Sbjct: 361 IFGNSTGGRRGAMIGA 376
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 155

A DNA sequence (GBSx0161) was identified in *S.agalactiae* <SEQ ID 517> which encodes the amino acid sequence <SEQ ID 518>. This protein is predicted to be transketolase, N-terminal subunit (tkt). Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3680(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB98676 GB:U67515 transketolase' [Methanococcus jannaschii]
 Identities = 106/269 (39%), Positives = 158/269 (58%), Gaps = 4/269 (1%)

Query: 11  LRRFATEIRLNTLETLNHLGFGHYGGSLSIVEALAVLYGDIMDINPEKFKESDRDYMVLS 70
           L + A ++R N ++ +     GH GGSLS  + +  LY  +M+ +P+   + DRD  VLS
Sbjct: 10  LEKIAKKVRYNIVKMVGLAKSGHPGGSLSATDIIVALYFKLMNYSPDNPYKKDRDRFVLS 69

Query: 71  KGHAGPALYSTLYLKGFFDKTFLHSLNTNGTKLPSHPDRNLTPGIDVTTGSLGQGISIAT 130
           KGHA PALY+ L   G  ++  L  L     KL  HP  + TPG+++ TGSLGQG S A
Sbjct: 70  KGHAAPALYAVLSELGIIEEEELWKLRRLEGKLQGHPSMD-TPGVEICTGSLGQGFSAAV 128

Query: 131 GIAYAQKIENSSYYTYTIVGDGELNEGQCWEAIQFAAHHQLHHLIVFVDDNKKQLDGLTA 190
           G+A   +++  + Y Y ++GDGE  EG  WEA   AAH++L +LI F+D NK Q+DG T
Sbjct: 129 GMALGCRLDKLNNYVYVLLGDGECQEGIVWEAAMAAAHYKLDNLIAFIDRNKLQIDGCTE 188

Query: 191 DICNPGDFVAKFEAFGFDAVRVKGDDIEAIDKAIKTFQDSNSVRPKCIVLDSIKGQGVKE 250
           D+ + GD  AKFEAFG+D   + G + E I   ++  +   + +PK I+  ++KG+GV
Sbjct: 189 DVMSLGDIKAKFEAFGWDVFEIDGHNFEEIINTVEKAKSMKNGKPKMIIAYTVKGKGVSF 248

Query: 251 LEELASNHHLRPDLQQKTMLERALISLRE 279
           +E   + H   P+ +Q   L++AL  L E
Sbjct: 249 MENNVAFHGKAPNEEQ---LKQALEELSE 274
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 519> which encodes the amino acid sequence <SEQ ID 520>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.75   Transmembrane   58 -  74 (  57 -  74)

----- Final Results -----
               bacterial membrane --- Certainty=0.1298(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9165> which encodes the amino acid sequence <SEQ ID 9166>. Analysis of this protein sequence reveals the following:

```
     Possible site: 54
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.75   Transmembrane   40 -  56 (  39 -  56)

----- Final Results -----
               bacterial membrane --- Certainty=0.130(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 82/246 (33%), Positives = 129/246 (52%), Gaps = 15/246 (6%)

Query: 18  IRLNTLETLNHLGFGHYGGSLSIVEALAVLYGDIMDINPEKFKE-SDRDYMVLSKGHAGP 76
           +R  +++ +     GH G  +        VL+   M+INP+  + S+RD  +LS GH
Sbjct: 82  VRTLSMDAIQAANSGHPGLPMGAAPMAYVLWNHFMNINPKTSRNWSNRDRFILSAGHGSA 141

Query: 77  ALYSTLYLKGF-FDKTFLHSLNTNGTKLPSHPDRNLTPGIDVTTGSLGQGISIATGIAYA 135
            LYS L+L G+      L +    G+K P HP+ N T G++ TTG LGQGI+ A G+A A
Sbjct: 142 MLYSLLHLAGYDLSVEDLKNFRQWGSKTPGHPEVNHTDGVEATTGPLGQGIANAVGMAMA 201

Query: 136 QK----------IENSSYYTYTIVGDGELNEGQCWEAIQFAAHHQLHHLIVFVDDNKKQL 185
           +           +   +YT+ + GDG+L EG   EA   A H +L  L++  D N   L
Sbjct: 202 EAHLAAKFNKPGFDIVDHYTFALNGDGDLMEGVSQEAASMAGHLKLGKLVLLYDSNDISL 261
```

```
Query: 186 DGLTADICNPGDFVAKFEAFGFDAVRVK-GDDIEAIDKAIKTFQDSNSVRPKCIVLDSIK 244
           DG T+ +    D   +FEA+G+  + VK G+D+E I  AI+  + + + +P  I + +I
Sbjct: 262 DGPTS-MAFTEDVKGRFEAYGWQHILVKDGNDLEEIAAAIEAAK-AETEKPTIIEVKTII 319

Query: 245 GQGVKE 250
           G G ++
Sbjct: 320 GFGAEK 325
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 156

A DNA sequence (GBSx0162) was identified in *S.agalactiae* <SEQ ID 521> which encodes the amino acid sequence <SEQ ID 522>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.27   Transmembrane   53 -  69 (  53 -  69)

----- Final Results -----
               bacterial membrane --- Certainty=0.1107(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9499> which encodes amino acid sequence <SEQ ID 9500> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB98674 GB:U67515 transketolase'' [Methanococcus jannaschii]
 Identities = 100/301 (33%), Positives = 171/301 (56%), Gaps = 7/301 (2%)

Query: 6   KEMRLVYRDFLLQANQENKQITVLEADLSSSMSTNALASEFGKRYINLGIMEAEMVGLAA 65
           K MR  Y + L++  ++ + + VL+ADLS S  T   A EF +R+ N G+ E  M+G+AA
Sbjct: 9   KGMRKGYGETLIELGKKYENLVVLDADLSGSTQTAMFAKEFPERFFNAGVAEQNMIGMAA 68

Query: 66  GLAIKGYKPYLHTFGPFASRRVFDQVFLSLGYSQLSATIIGSDAGISAEMNGGTHMPFEE 125
           GLA  G   +  +F  FAS R ++ +   + Y +L+  I+ + AGI+   +G +H   E+
Sbjct: 69  GLATTGKIVFASSFSMFASGRAWEIIRNLVAYPKLNVKIVATHAGITVGEDGASHQMCED 128

Query: 126 LGLLRLIPKATIFEVSDDIQFEAILKQTLSIDGLKYIRTIRKAPTAVYEGRE----DFSK 181
           + ++R IP   +   +D   + +++      G  Y+R  R+    +YE  E    +  K
Sbjct: 129 IAIMRAIPNMVVIAPTDYYHTKNVIRTIAEYKGPVYVRMPRRDTEIIYENEEEATFEIGK 188

Query: 182 GFIQLRQGKDITLVASGIMVSRAIEAADYLKELGIEASVIDLFKIKPLPEELKPLLIDQS 241
           G I L  G+D+T++A+G  V  A+ A + LKE GI A ++++  IKP+ EE+     D
Sbjct: 189 GKI-LVDGEDLTIIATGEEVPEALRAGEILKENGISAEIVEMATIKPIDEEIIKKSKD-F 246

Query: 242 IVTIENHNRIGGIGSALCEWL-SMEKDTTVSRMGIDERFGQVGQMEYLLEEYGLAVKDIVQ 301
           +VT+E+H+ IGG+G A+ E + S   +  + R+GI++ FG+ G+ + LL+ YGL  + I +
Sbjct: 247 VVTVEDHSIIGGLGGAVAEVIASNGLNKKLLRIGINDVFGRSGKADELLKYYGLDGESIAK 307
```

There is also homology to SEQ ID 520.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 157

A DNA sequence (GBSx0163) was identified in *S.agalactiae* <SEQ ID 523> which encodes the amino acid sequence <SEQ ID 524>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2517(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 158

A DNA sequence (GBSx0164) was identified in *S.agalactiae* <SEQ ID 525> which encodes the amino acid sequence <SEQ ID 526>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -6.42   Transmembrane  119 - 135 ( 114 - 145)
   INTEGRAL    Likelihood = -5.10   Transmembrane   33 -  49 (  32 -  50)
   INTEGRAL    Likelihood = -4.30   Transmembrane   94 - 110 (  94 - 111)
   INTEGRAL    Likelihood = -3.66   Transmembrane   67 -  83 (  60 -  83)

----- Final Results -----
              bacterial membrane --- Certainty=0.3569(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8503> and protein <SEQ ID 8504> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 4
SRCFLG: 0
McG: Length of UR:   22
     Peak Value of UR:   2.96
     Net Charge of CR: 2
McG: Discrim Score:     10.55
GvH: Signal Score (-7.5): -4.31
     Possible site: 22
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program   count: 6 value:  -6.42 threshold:  0.0
   INTEGRAL    Likelihood = -6.42   Transmembrane  154 - 170 ( 149 - 180)
   INTEGRAL    Likelihood = -5.10   Transmembrane   68 -  84 (  67 -  85)
   INTEGRAL    Likelihood = -5.04   Transmembrane    6 -  22 (   2 -  24)
   INTEGRAL    Likelihood = -4.30   Transmembrane  129 - 145 ( 129 - 146)
   INTEGRAL    Likelihood = -3.66   Transmembrane  102 - 118 (  95 - 118)
   INTEGRAL    Likelihood = -3.56   Transmembrane   29 -  45 (  29 -  46)
   PERIPHERAL  Likelihood =  0.79     285
 modified ALOM score:   1.78
icm1 HYPID: 7  CFP: 0.357

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.3569(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01868(391 - 1575 of 1938)
GP|9946413|gb|AAG03934.1|AE004491_1|AE004491(5   -   434   of   434)   hypothetical   protein
{Pseudomonas aeruginosa}
%Match = 8.1
%Identity = 26.1  %Similarity = 48.6
Matches = 105  Mismatches = 192  Conservative Sub.s = 91

171       201       231       261       291       321       351       381
DTTVSRMGIDERFGQVGQMEYLLEEYGLAVKDIVQHCKSIYKS*QKGNIGVAFLLFSEIFKFCISILWYFILTKNKGVVV

                                                                          M

411       441       471       480       507       537       567       597
MRAWKGIVLILSSIVVTLVAWQNAGLSEFVV-------PGLALTSL-SLTFLLSTKFRILESYFQGIENMYFYHKVMAVF
   |  :   |::  :||     | |  :||        |:|  :| ||  ||: :   ||  : |:: ||  ||   ::
KLLWGVLAAALAAWGLTLAVDPPASLDIWVVRKQAILLTGVASFALMSLIMLLAVRPVWLEKPLDGLDRMYRLHKWAGIL
                 20        30        40        50        60        70        80

627       657                              687       717       747       777
SMILLLLHKIGLGQGGHGSEF-------------------AKTIGSAGLYLFLSIVFVAYFGNFLKYEIWRFIHRFVYL
:::| |||  :    |   :                       || :|    :::  :::|  :   : | :||::|: : |
AIVLGLLHYLLELAGPWLAGIVGKPVKGPRVETFLDVFRGSAKELGEWSAWILGGMLLVTLW-QRFPYHLWRYVHKALAL
                100       110       120       130       140       150       160

807       837       867       897       924       951       981      1011
AYILGLVHTFMILGDRILGNTLLSLIVLGYAVIGVISGFYIIFLYSRM-RFRR-VGYVQKVTHLNHDTTEIEIAMKRPYR
 |::   |: ::|      :     :|   |::|      :: |  |: | ||  | |  |         |:   ::  :
VYLVLAFHS-VVLAPASYWSQPAGWLVAACALLGSACA--LLSLSGRIGRTRRHAGVVTAVERHGESLLEVTCRLQGDWS
        170       180       190       200       210       220       230

1041      1071      1101      1125      1155      1185      1215      1242
YDYGQFTFFKIYQAGFESAAHPFSISGGHDRV--IFLTVKASGDYTKSIYKQLKVGTKIALDRAYGHMLFDKD-KKEQVW
:  ||| |           ||||:|:       : :::|| ||||: :   |:|| :: ::  ||   | :     |||
HRAGQFAF---LTCDRLEGAHPFTIASADRGCGEVRFSIKALGDYTRRLQDNLEVGARVEVEGPYGCFDFRRGLAGRQVW
             250       260       270       280       290       300       310

1272      1293      1323      1353      1383      1413      1443      1461
IAGGIGITPFISFI---RENSILTKRVDFFYTFSNQDNLIYQDMLESYAKANPNFKLHLNNSSLKGRLDFSQ----SVFE
:| |||:||||:::   :        |:: |   |    ::   |    :  |:  ||:  |  :|:   :|       |
VAAGIGVTPFIAWLESLQAAPESAPSVELHYCVRNSQEALFAGRLRELCEHLPSVTLHIRYSDEQGKPQAAQLGVLKSAE
             330       340       350       360       370       380       390

1488      1518      1548      1575      1605      1635      1665      1695
GQ-PTIFMCGPTSMTSTYAKVFRQKDAKSRLVY-EGFSFRDSWLSIFLLKTFDKVYSNLIK*EGL*DKPTFSWF*ECQS*
|: |::: |||  :  :  : :|::    || : | |  |
GRWPSVWFCGPQGLADSLRRDLRRQGMPLRLFHQEAFRMR
             410       420       430
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 159

A DNA sequence (GBSx0165) was identified in *S.agalactiae* <SEQ ID 527> which encodes the amino acid sequence <SEQ ID 528>. This protein is predicted to be 30S ribosomal protein S15 (rpsO). Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4074(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB13541 GB:Z99112 ribosomal protein S15 (BS18) [Bacillus subtilis]
 Identities = 55/89 (61%), Positives = 71/89 (78%)

Query: 1  MAISKEKKNEIIAQYARHEGDTGSVEVQVAVLTWEINHLNDHIKQHKKDHATYRGLMKKI 60
          MAI++E+KN++I ++  HE DTGS EVQ+A+LT  IN+LN+H++ HKKDH + RGL+K +
Sbjct: 1  MAITQERKNQLINEFKTHESDTGSPEVQIAILTDSINNLNEHLRTHKKDHHSRRGLLKMV 60

Query: 61 GHRRNLLAYLRRTDVNRYRELIQSLGLRR 89
          G RRNLL YLR  DV RYRELI  LGLRR
Sbjct: 61 GKRRNLLTYLRNKDVTRYRELINKLGLRR 89
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 529> which encodes the amino acid sequence <SEQ ID 530>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3746(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 88/89 (98%), Positives = 88/89 (98%)

Query: 1  MAISKEKKNEIIAQYARHEGDTGSVEVQVAVLTWEINHLNDHIKQHKKDHATYRGLMKKI 60
          MAISKEKKNEIIAQYARHEGDTGSVEVQVAVLTWEINHLN HIKQHKKDHATYRGLMKKI
Sbjct: 1  MAISKEKKNEIIAQYARHEGDTGSVEVQVAVLTWEINHLNSHIKQHKKDHATYRGLMKKI 60

Query: 61 GHRRNLLAYLRRTDVNRYRELIQSLGLRR 89
          GHRRNLLAYLRRTDVNRYRELIQSLGLRR
Sbjct: 61 GHRRNLLAYLRRTDVNRYRELIQSLGLRR 89
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 160

A DNA sequence (GBSx0166) was identified in *S.agalactiae* <SEQ ID 531> which encodes the amino acid sequence <SEQ ID 532>. This protein is predicted to be polyribonucleotide nucleotidyltransferase (pnp). Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.64   Transmembrane  448 - 464 ( 448 - 464)

----- Final Results -----
               bacterial membrane --- Certainty=0.1256(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9501> which encodes amino acid sequence <SEQ ID 9502> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC43595 GB:U29668 polynucleotide phosphorylase [Bacillus subtilis]
 Identities = 428/694 (61%), Positives = 532/694 (75%), Gaps = 4/694 (0%)

Query: 7   KQVFEMIFAGKKLVVETGQVAKQANGSVVVRYGDSTVLTAAVMSKKMSTGDFFPLQVNYE 66
```

```
            K VF + +AG+ L VETGQ+AKQANG+V++RYGD+ VL+ A  SK+    DFFPL VNYE
Sbjct: 5    KHVFTIDWAGRTLTVETGQLAKQANGAVMIRYGDTAVLSTATASKEPKPLDFFPLTVNYE 64

Query: 67   EKMYAAGKFPGGFNKREGRPSTDATLTARLIDRPIRPMFAEGFRNEVQVINTVLSFDENA 126
            E++YA GK PGGF KREGRPS  A L +RLIDRPIRP+FA+GFRNEVQVI+ V+S D+N
Sbjct: 65   ERLYAVGKIPGGFIKREGRPSEKAVLASRLIDRPIRPLFADGFRNEVQVISIVMSVDQNC 124

Query: 127  SAPMAAMFGSSLALSISDIPFNGPIAGVQVAYVDGNFIINPTAQEQEASALELTVAGTKE 186
            S+ MAAMFGSSLALS+SDIPF GPIAGV V  +D  FIINPT  + E S + L VAGTK+
Sbjct: 125  SSEMAAMFGSSLALSVSDIPFEGPIAGVTVGRIDDQFIINPTVDQLEKSDINLVVAGTKD 184

Query: 187  AINMVESGAKELSEEIMLEALLKGHEAVCELIAFQEEIVTAIGKEKAEVELLQVDPELQA 246
            AINMVE+GA E+ EEIMLEA++ GHE +  LIAFQEEIV A+GKEK+E++L ++D EL
Sbjct: 185  AINMVEAGADEVPEEIMLEAIMFGHEEIKRLIAFQEEIVAAVGKEKSEIKLFEIDEELNE 244

Query: 247  EIIATHNIALQAAVQVEEKKAREAATEAVKEVVIGEYEARYAEHEEYDRIMRDVAEILEQ 306
            ++ A     L  A+QV EK ARE A   VK  V+ ++E    EH+E    ++ V +IL +
Sbjct: 245  KVKALAEEDLLKAIQVHEKHAREDAINEVKNAVVAKFEDE--EHDE--DTIKQVKQILSK 300

Query: 307  MEHAEVRRLITEDKIRPDGRRVDEIRPLDAEIDFLPQVHGSGLFTRGQTQALSVLTLAPM 366
            +   EVRRLITE+K+RPDGR VD+IRPL +E+  LP+ HGSGLFTRGQTQALSV TL  +
Sbjct: 301  LVKNEVRRLITEEKVRPDGRGVDQIRPLSSEVGLLPRTHGSGLFTRGQTQALSVCTLGAL 360

Query: 367  GEAQIIDGLTPEYKKRFMHHYNFPQYSVGETGRYGAAGRREIGHGALGERALEQVLPRLE 426
            G+ QI+DGL  E  KRFMHHYNFPQ+SVGETG    GRREIGHGALGERALE V+P  +
Sbjct: 361  GDVQILDGLGVEESKRFMHHYNFPQFSVGETGPMRGPGRREIGHGALGERALEPVIPSEK 420

Query: 427  EFPYAIRLVAEVLESNGSSSQASICAGTLALMAGGVPIKAPVAGIAMGLISDGTNYTVLT 486
            +FPY +RLV+EVLESNGS+SQASICA TLA+M  GVPIKAPVAGIAMGL+  G +YTVLT
Sbjct: 421  DFPYTVRLVSEVLESNGSTSQASICASTLAMMDAGVPIKAPVAGIAMGLVKSGEHYTVLT 480

Query: 487  DIQGLEDHFGDMDFKVAGTREGITALQMDIKIEGITPQILEEALAQAKKARFEILDVLHG 546
            DIQG+ED  GDMDFKVAGT +G+TALQMDIKIEG++ +ILEEAL QAKK R EIL+ +
Sbjct: 481  DIQGMEDALGDMDFKVAGTEKGVTALQMDIKIEGLSREILEEALQQAKKGRMEILNSMLA 540

Query: 547  AIAEPRPQLAPTAPKIDMIKIDVDKIKVVIGKGGETIDKIIAETGVKIDIDEEGNVSIFS 606
             ++E R +L+  APKI  + I+ DKI+ VIG  G+ I+KII ETGVKIDI+++G + I S
Sbjct: 541  TLSESRKELSRYAPKILTMTINPDKIRDVIGPSGKQINKIIEETGVKIDIEQDGTIFISS 600

Query: 607  SDQAAIDRTKDIIASLVREAKVGEVYHAKVVRIEKFGAFVNLFDKTDALVHISEIAWTRT 666
            +D++   + K II  LVRE +VG++Y  KV RIEKFGAFV +F   D LVHISE+A  R
Sbjct: 601  TDESGNQKAKKIIEDLVREVEVGQLYLGKVKRIEKFGAFVEIFSGKDGLVHISELALERV 660

Query: 667  ANVADVLEIGEEVDVKVIKIDDKGRVDASMKALL 700
              V DV++IG+E+ VKV +ID +GRV+ S KA+L
Sbjct: 661  GKVEDVVKIGDEILVKVTEIDKQGRVNLSRKAVL 694
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 533> which encodes the amino acid sequence <SEQ ID 534>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.64   Transmembrane  444 - 460 ( 444 - 460)

----- Final Results -----
               bacterial membrane --- Certainty=0.1256(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 631/708 (89%), Positives = 664/708 (93%), Gaps = 2/708 (0%)

Query: 5    MSKQVFEMIFAGKKLVVETGQVAKQANGSVVVRYGDSTVLTAAVMSKKMSTGDFFPLQVN 64
            MSKQ F   FAGK LVVE GQVAKQANG+ VVRYGDSTVLTAAVMSKKM+TGDFFPLQVN
Sbjct: 1    MSKQTFTTTFAGKPLVVEVGQVAKQANGATVVRYGDSTVLTAAVMSKKMATGDFFPLQVN 60
```

```
Query: 65  YEEKMYAAGKFPGGFNKREGRPSTDATLTARLIDRPIRPMFAEGFRNEVQVINTVLSFDE 124
           YEEKMYAAGKFPGGF KREGRPSTDATLTARLIDRPIRPMFAEGFRNEVQVINTVLS+DE
Sbjct: 61  YEEKMYAAGKFPGGFMKREGRPSTDATLTARLIDRPIRPMFAEGFRNEVQVINTVLSYDE 120

Query: 125 NASAPMAAMFGSSLALSISDIPFNGPIAGVQVAYVDGNFIINPTAQEQEASALELTVAGT 184
           NASAPMAAMFGSSLALSISDIPFNGPIAGVQV Y+DG FIINP  ++ EAS LELTVAG+
Sbjct: 121 NASAPMAAMFGSSLALSISDIPFNGPIAGVQVGYIDGEFIINPDKEQMEASLLELTVAGS 180

Query: 185 KEAINMVESGAKELSEEIMLEALLKGHEAVCELIAFQEEIVTAIGKEKAEVELLQVDPEL 244
           KEAINMVESGAKELSE+IMLEALLKGH+A+ ELIAFQE+IV  +GKEKAEVELLQVD +L
Sbjct: 181 KEAINMVESGAKELSEDIMLEALLKGHQAIQELIAFQEQIVAVVGKEKAEVELLQVDVDL 240

Query: 245 QAEIIATHNIALQAAVQVEEKKAREAATEAVKEVVIGEYEARYAEHEEYDRIMRDVAEIL 304
           QA+I+A +N  LQ AVQVEEKKAREAATEAVKE+V  EYE RYAE E    IMRDVAEIL
Sbjct: 241 QADIVAKYNAQLQKAVQVEEKKAREAATEAVKEMVKAEYEERYAEDENLATIMRDVAEIL 300

Query: 305 EQMEHAEVRRLITEDKIRPDGRRVDEIRPLDAEIDFLPQVHGSGLFTRGQTQALSVLTLA 364
           EQMEHAEVRRLITEDKIRPDGR++DEIRPLDA +DFLP+VHGSGLFTRGQTQALSVLTLA
Sbjct: 301 EQMEHAEVRRLITEDKIRPDGRKIDEIRPLDAVVDFLPKVHGSGLFTRGQTQALSVLTLA 360

Query: 365 PMGEAQIIDGLTPEYKKRFMHHYNFPQYSVGETGRYGAAGRREIGHGALGERALEQVLPR 424
           PMGE QIIDGL PEYKKRF+HHYNFPQYSVGETGRYGAAGRREIGHGALGERALEQVLP
Sbjct: 361 PMGETQIIDGLAPEYKKRFLHHYNFPQYSVGETGRYGAAGRREIGHGALGERALEQVLPS 420

Query: 425 LEEFPYAIRLVAEVLESNGSSSQASICAGTLALMAGGVPIKAPVAGIAMGLISDGTNYTV 484
           LEEFPYAIRLVAEVLESNGSSSQASICAGTLALMAGGVPIKAPVAGIAMGLISDGTNYTV
Sbjct: 421 LEEFPYAIRLVAEVLESNGSSSQASICAGTLALMAGGVPIKAPVAGIAMGLISDGTNYTV 480

Query: 485 LTDIQGLEDHFGDMDFKVAGTREGITALQMDIKIEGITPQILEEALAQAKKARFEILDVL 544
           LTDIQGLEDHFGDMDFKVAGTREGITALQMDIKI GITPQILEEALAQAKKARFEILDV+
Sbjct: 481 LTDIQGLEDHFGDMDFKVAGTREGITALQMDIKIAGITPQILEEALAQAKKARFEILDVI 540

Query: 545 HGAIAEPRPQLAPTAPKIDMIKIDVDKIKVVIGKGGETIDKIIAETGVKIDIDEEGNVSI 604
              IAEPRP+LAPTAPKID IKIDVDKIKVVIGKGGETIDKIIAETGVKIDID+EGNVSI
Sbjct: 541 EATIAEPRPELAPTAPKIDTIKIDVDKIKVVIGKGGETIDKIIAETGVKIDIDDEGNVSI 600

Query: 605 FSSDQAAIDRTKDIIASLVREAKVGEVYHAKVVRIEKFGAFVNLFDKTDALVHISEIAWT 664
           +SSDQAAIDRTK+IIA LVREAKVGEVYHAKVVRIEKFGAFVNLFDKTDALVHISEIAWT
Sbjct: 601 YSSDQAAIDRTKEIIAGLVREAKVGEVYHAKVVRIEKFGAFVNLFDKTDALVHISEIAWT 660

Query: 665 RTANVADVLEIGEEVDVKVIKIDDKGRVDASMKALLPRPPKADNPKKE 712
           RT NV+DVLE+GE+VDVKVIKID+KGRVDASMKAL+PRPPK +  KKE
Sbjct: 661 RTTNVSDVLEVGEDVDVKVIKIDEKGRVDASMKALIPRPPKPE--KKE 706
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 161

A DNA sequence (GBSx0167) was identified in *S.agalactiae* <SEQ ID 535> which encodes the amino acid sequence <SEQ ID 536>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1293(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 537> which encodes the amino acid sequence <SEQ ID 538>. Analysis of this protein sequence reveals the following:

```
     Possible site: 38

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.43   Transmembrane   83 -  99 (  83 -  99)

----- Final Results -----
               bacterial membrane --- Certainty=0.1171(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 172/248 (69%), Positives = 211/248 (84%)

Query: 1   MTSTNELDIRLRAFINAPDNFLDSIGLVNALHHSTVWASKEPYAIQVDGQEVVPVFTDIT 60
           MT +NELDIRLRAFINAPDNFLDS+ LVNA H+  VWA+KEPY I+V+G +V PVFTD
Sbjct: 1   MTKSNELDIRLRAFINAPDNFLDSLALVNAFHNFPVWAAKEPYVIEVEGVKVTPVFTDKE 60

Query: 61  DLNHFKEEQESARDMFWESRRSLDVLDEAISHGLAGLVYNLKKEGDFGNSTIFYCEDMVQ 120
           D+  FKEEQ+SA+  +W  R +L VL+E I+ G AGL++NLKK+GDFGNSTIF   DM+Q
Sbjct: 61  DMARFKEEQKSAQSQYWLERSALAVLEEVITSGAAGLIFNLKKKGDFGNSTIFKSSDMIQ 120

Query: 121 FMNNYTTILNQLLNEDNIVADIMDKTYLVPAFVHPREEGSFDRLFPTMSTPEGKSYVPVF 180
           FMN+YTT+LN L+++DN+ AD M+K YLVPAFV+P++   +DRLFPTMSTPEGKSYVP F
Sbjct: 121 FMNHYTTVLNTLMSDDNVAADTMEKVYLVPAFVYPKDNNHYDRLFPTMSTPEGKSYVPAF 180

Query: 181 SNLLSFEKWYNHNDFGGAFRKAQGVILAWTIDDIYKPRNGENEIDDTFGVAINPFDEQQV 240
           SNL SF KWYN +DFGG FRKA+GVIL WTIDDIY+PRNGENE+D+TFGVAINPFD+QQ+
Sbjct: 181 SNLQSFAKWYNQDDFGGLFRKAEGVILTWTIDDIYQPRNGENELDETFGVAINPFDDQQI 240

Query: 241 LVDWSDVE 248
           LVDWS+++
Sbjct: 241 LVDWSELD 248
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 162

A DNA sequence (GBSx0168) was identified in *S.agalactiae* <SEQ ID 539> which encodes the amino acid sequence <SEQ ID 540>. This protein is predicted to be serine acetyltransferase (cysE). Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.02   Transmembrane  150 - 166 ( 147 - 168)

----- Final Results -----
               bacterial membrane --- Certainty=0.1808(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9503> which encodes amino acid sequence <SEQ ID 9504> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB71304 GB:AJ130879 serine acetyltransferase [Clostridium
           sticklandii]
 Identities = 92/169 (54%), Positives = 125/169 (73%)

Query: 9   KESIAIVKEQDPAARSSLEVILTYPGIKALAAHRLSHFLWNHNFKLLARMHSQFWRFWTQ 68
```

```
           KE+I + +E+DPAA+ ++ +++  PGI A+  HR++H L+N     +AR+ SQ  RF T
Sbjct: 20  KETIEVAREKDPAAKGAINILVNTPGIHAIMFHRVAHSLYNRKHFFIARLISQISRFLTG 79

Query: 69  IEIHPGATISEGVFIDHGSGLVIGETAIVEKGAMLYHGVTLGGTGKDKGKRHPTIRKGAL 128
           IEIHPGA I    FIDHG G+VIGETA +    ML+H VTLGGTGKDKGKRHPT+    +
Sbjct: 80  IEIHPGAQIGRRFFIDHGMGVVIGETAEIGDDVMLFHQVTLGGTGKDKGKRHPTVENNVI 139

Query: 129 ISAHSQIIGPIEVGENAKVGAAAVVLADVPADVTVVGVPAKVVRVHGQK 177
           ISA  +++GPI +GEN+K+GA AVVL D+P + T VG+PAKVVR++G+K
Sbjct: 140 ISAGVKVLGPIVIGENSKIGANAVVLHDIPKNATAVGIPAKVVRLNGEK 188
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 541> which encodes the amino acid sequence <SEQ ID 542>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0141(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 162/193 (83%), Positives = 178/193 (91%)

Query: 5   MGWWKESIAIVKEQDPAARSSLEVILTYPGIKALAAHRLSHFLWNHNFKLLARMHSQFWR 64
           MGWWKESIAIVK  DPAAR+SLEVILTYPGIKALAAHRLSHFLW H+FKLLARMHSQFWR
Sbjct: 1   MGWWKESIAIVKALDPAARNSLEVILTYPGIKALAAHRLSHFLWRHHFKLLARMHSQFWR 60

Query: 65  FWTQIEIHPGATISEGVFIDHGSGLVIGETAIVEKGAMLYHGVTLGGTGKDKGKRHPTIR 124
           FWTQIEIHPGA I+ GVFIDHG+GLVIGETAIVEKG MLYHGVTLGGTGKD GKRHPT+R
Sbjct: 61  FWTQIEIHPGAQIAPGVFIDHGAGLVIGETAIVEKGVMLYHGVTLGGTGKDCGKRHPTVR 120

Query: 125 KGALISAHSQIIGPIEVGENAKVGAAAVVLADVPADVTVVGVPAKVVRVHGQKDDLQIRS 184
           +GALISAH+Q+IGPI++G NAKVGAAAVVL+DVP DVTVVGVPAK+VRVHGQKD+ QI+S
Sbjct: 121 QGALISAHAQVIGPIDIGANAKVGAAAVVLSDVPEDVTVVGVPAKIVRVHGQKDNRQIQS 180

Query: 185 IEHDREESYYSSK 197
           ++  RE SY  SK
Sbjct: 181 LQKQREVSYQLSK 193
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 163

A DNA sequence (GBSx0169) was identified in *S.agalactiae* <SEQ ID 543> which encodes the amino acid sequence <SEQ ID 544>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> May be a lipoprotein
   INTEGRAL    Likelihood = -5.89   Transmembrane   32 -  48 (  29 -  49)

----- Final Results -----
               bacterial membrane --- Certainty=0.3357(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 164

A DNA sequence (GBSx0170) was identified in *S.agalactiae* <SEQ ID 545> which encodes the amino acid sequence <SEQ ID 546>. This protein is predicted to be cysteinyl-tRNA synthetase (cysS). Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2227(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB11870 GB:Z99104 cysteinyl-tRNA synthetase [Bacillus subtilis]
 Identities = 246/465 (52%), Positives = 322/465 (68%), Gaps = 23/465 (4%)

Query: 2   IKIYDTMTRSLQDFIPLNEGKVNMYVCGPTVYNYIHIGNARSVVAFDTIRRYFEYCGYQV 61
           I +Y+T+TR  + F+PL EGKV MYVCGPTVYNYIHIGNAR  + +DT+R Y EY GY V
Sbjct: 3   ITLYNTLTRQKETFVPLEEGKVKMYVCGPTVYNYIHIGNARPAIVYDTVRNYLEYKGYDV 62

Query: 62  NYISNFTDVDDKIIKGAAEAGMDTKSFSDKFISAFMEDVAALGVKPATKNPRVIDYMDEI 121
            Y+SNFTDVDDK+IK A E G D  + S++FI A+ EDV ALG + A  +PRV++ MD I
Sbjct: 63  QYVSNFTDVDDKLIKAANELGEDVPTISERFIKAYFEDVGALGCRKADLHPRVMENMDAI 122

Query: 122 IDFVKVLVDKEFAYEANGDVYFRVSKSHHYAKLANKTLEDLEIGASGRVDGEGEIKENPL 181
           I+FV  LV K +AYE+ GDVYF+       Y KL+ +++++L  GA  RV   GE KE+ L
Sbjct: 123 IEFVDQLVKKGYAYESEGDVYFKTRAFEGYGKLSQQSIDELRSGARIRV---GEKKEDAL 179

Query: 182 DFALWKSAKSGEVSWESPWGKGRPGWHIECSVMATEILGDTIDIHGGGADLEFPHHTNEI 241
           DFALWK+AK GE+SW+SPWGKGRPGWHIECS M  + LGD IDIH GG DL FPHH NEI
Sbjct: 180 DFALWKAAKEGEISWDSPWGKGRPGWHIECSAMVKKYLGDQIDIHAGGQDLTFPHHENEI 239

Query: 242 AQSEAKTGKTFANYWMHNGFVNVDNEKMSKSLGNFITVHDMLKSVDGQVIRFFLATQQYR 301
           AQSEA TGKTFA YW+HNG++N+DNEKMSKSLGNF+ VHD++K  D Q++RFF+ +  YR
Sbjct: 240 AQSEALTGKTFAKYWLHNGYINIDNEKMSKSLGNFVLVHDIIKQHDPQLLRFFMLSVHYR 299

Query: 302 KPVNFTEKAVHDAEVNLKYLKNTF----------NLPIQENANDEELEQFVKAFQGAMD 350
            P+N++E+ + + +     LK  +           NL   ++   E++E+  KAF+  MD
Sbjct: 300 HPINYSEELLENTKSAFSRLKTAYSNLQHRLNSSTNLTEDDDQWLEKVEEHRKAFEEEMD 359

Query: 351 DDFNTANGITVIFEMAKWIN--------SGHYTSRVKETFAELLEIFGI-VFQEEVLDAD 401
           DDFNTAN I+V+F++AK  N        + H  +   E F  ++ + G  + ++E+LD +
Sbjct: 360 DDFNTANAISVLFDLAKHANYYLQKDHTADHVITAFIEMFDRIVSVLGFSLGEQELLDQE 419

Query: 402 IESLIEQRQEARANRDFATADRIRDELAKQGIKLLDTKDGVRWTR 446
           IE LIE+R EAR NRDFA +D+IRD+L    I L DT  G RW R
Sbjct: 420 IEDLIEKRNEARRNRDFALSDQIRDQLKSMNIILEDTAQGTRWKR 464
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 547> which encodes the amino acid sequence <SEQ ID 548>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.1765(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 357/447 (79%), Positives = 401/447 (88%)

Query: 1   MIKIYDTMTRSLQDFIPLNEGKVNMYVCGPTVYNYIHIGNARSVVAFDTIRRYFEYCGYQ 60
           MIKIYDTMTRSL+ F+PL E  VN+YVCGPTVYNYIHIGNARS VAFDTIRRYFEY GYQ
Sbjct: 1   MIKIYDTMTRSLRKFVPLTENTVNIYVCGPTVYNYIHIGNARSAVAFDTIRRYFEYTGYQ 60

Query: 61  VNYISNFTDVDDKIIKGAAEAGMDTKSFSDKFISAFMEDVAALGVKPATKNPRVIDYMDE 120
           VNYISNFTDVDDKIIK A +AG+  K  SD+FI+AF+ED  ALGVKPAT+NPRV+DY+ E
Sbjct: 61  VNYISNFTDVDDKIIKAATQAGVSPKELSDRFIAAFIEDTKALGVKPATQNPRVMDYIAE 120

Query: 121 IIDFVKVLVDKEFAYEANGDVYFRVSKSHHYAKLANKTLEDLEIGASGRVDGEGEIKENP 180
           II FV+ L++K+FAYEA+GDVYFRV KS HYAKLANKTL +LE+GASGR D E  +KENP
Sbjct: 121 IISFVESLIEKDFAYEADGDVYFRVEKSEHYAKLANKTLSELEVGASGRTDAETALKENP 180

Query: 181 LDFALWKSAKSGEVSWESPWGKGRPGWHIECSVMATEILGDTIDIHGGGADLEFPHHTNE 240
           LDFALWKSAK+GEVSW+SPWG GRPGWHIECSVMATEILGDTIDIHGGGADLEFPHHTNE
Sbjct: 181 LDFALWKSAKAGEVSWDSPWGFGRPGWHIECSVMATEILGDTIDIHGGGADLEFPHHTNE 240

Query: 241 IAQSEAKTGKTFANYWMHNGFVNVDNEKMSKSLGNFITVHDMLKSVDGQVIRFFLATQQY 300
           IAQSEAKTGKTFANYWMHNGFV VDNEKMSKSLGNF+TVHDML++VDGQV+RFFLATQQY
Sbjct: 241 IAQSEAKTGKTFANYWMHNGFVTVDNEKMSKSLGNFVTVHDMLQTVDGQVLRFFLATQQY 300

Query: 301 RKPVNFTEKAVHDAEVNLKYLKNTFNLPIQENANDEELEQFVKAFQGAMDDDFNTANGIT 360
           RKP+NFTEK +HDAE+NLKYLKNT   P+ E A+++EL+QFV AFQ AMDDDFNTANGIT
Sbjct: 301 RKPINFTEKTIHDAEINLKYLKNTLQQPLTETADEQELKQFVIAFQDAMDDDFNTANGIT 360

Query: 361 VIFEMAKWINSGHYTSRVKETFAELLEIFGIVFQEEVLDADIESLIEQRQEARANRDFAT 420
           V+F+MAKWINSG YT  VK  F ++L +FGI+F+EEVL+ DIE+LI +RQEARANRDFAT
Sbjct: 361 VVFDMAKWINSGSYTEPVKSAFEKMLAVFGIIFEEEVLEVDIEALIAKRQEARANRDFAT 420

Query: 421 ADRIRDELAKQGIKLLDTKDGVRWTRD 447
           AD IRD+LA QGIKLLDTKDGVRW RD
Sbjct: 421 ADAIRDQLAVQGIKLLDTKDGVRWLRD 447
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 165

A DNA sequence (GBSx0171) was identified in *S.agalactiae* <SEQ ID 549> which encodes the amino acid sequence <SEQ ID 550>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0259(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9505> which encodes amino acid sequence <SEQ ID 9506> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB11871 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 58/122 (47%), Positives = 87/122 (70%)

Query: 3   DVRLINGIALAFEGDAVYSLYIRRHLIMQGFTKPNQLHRKATQYVSANAQALLINAMLEE 62
           D + +NG+ALA+ GDA++ +Y+R HL+ QGFTKPN LH+K++ VSA +QA ++  +  +
Sbjct: 9   DSKQLNGLALAYIGDAIFEVYVRHHLLKQGFTKPNDLHKKSSRIVSAKSQAEILFFLQNQ 68

Query: 63  NILTDEEQLIYKRGRNANSHTKAKNADIITYRMSTGFEALMGYLDMTGQIKRLETLIQWC 122
```

```
           +  T+EE+ + KRGRNA S T  KN D+ TYR ST FEAL+GYL +  + +RL  L+
Sbjct: 69  SFFTEEEEAVLKRGRNAKSGTTPKNTDVQTYRYSTAFEALLGYLFLEKKEERLSQLVAEA 128

Query: 123 IE 124
           I+
Sbjct: 129 IQ 130
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 551> which encodes the amino acid sequence <SEQ ID 552>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 99/127 (77%), Positives = 111/127 (86%)

Query: 2   IDVRLINGIALAFEGDAVYSLYIRRHLIMQGFTKPNQLHRKATQYVSANAQALLINAMLE 61
           +DV LINGIALAFEGDAVYS Y+RRHLI QG TKP+QLHR AT+YVSA AQA LI AMLE
Sbjct: 5   VDVNLINGIALAFEGDAVYSYYVRRHLIFQGKTKPSQLHRLATRYVSAKAQANLIQAMLE 64

Query: 62  ENILTDEEQLIYKRGRNANSHTKAKNADIITYRMSTGFEALMGYLDMTGQIKRLETLIQW 121
             +LT++E+ IYKRGRN NSHTKAKNADIITYRMSTGFEA+MGYLDM GQ +RLE LI+W
Sbjct: 65  AQLLTEKEEDIYKRGRNTNSHTKAKNADIITYRMSTGFEAIMGYLDMMGQKERLEELIRW 124

Query: 122 CIETIEK 128
           CIE +EK
Sbjct: 125 CIEYVEK 131
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 166

A DNA sequence (GBSx0172) was identified in *S.agalactiae* <SEQ ID 553> which encodes the amino acid sequence <SEQ ID 554>. This protein is predicted to be spoU rRNA methylase family protein. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1478(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB11872 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 113/244 (46%), Positives = 163/244 (66%), Gaps = 6/244 (2%)

Query: 11  ESSDLVYGLHAVTESLRANTG-NKLYLQDDLRGKNVDKVKALATEKKVSISWTPKKTLSD 69
           +  D V G +AV E+L+++     KL++ ++         +V  LA ++ ++I + P+K L
Sbjct: 3   QQHDYVIGKNAVIETLKSDRKLYKLWMAENTVKGQAQQVIELAKKQGITIQYVPRKKLDQ 62

Query: 70  MTNGGVHQGFVLKVSEFAYADLSEIMTKAENE-ENPLILILDGLTDPHNLGSILRTADAT 128
           M  G  HQG V +V+ + YA+L ++   AE + E P  LILD L DPHNLGSI+RTADA
Sbjct: 63  MVTGQ-HQGVVAQVAAYEYAELDDLYKAAEEKNEQPFFLILDELEDPHNLGSIMRTADAV 121
```

```
Query: 129 NVTGIIIPKHRSVGVTPVVSKTSTGAVEHVPIARVTNLSQTLDTLKDKEFWIFGTDMNGT 188
              GI+IPK R+VG+T  V+K STGA+EH+P+ARVTNL++TL+ +K++  W+ GTD +
Sbjct: 122 GAHGIVIPKRRAVGLTTTVAKASTGAIEHIPVARVTNLARTLEEMKERGIWVVGTDASAR 181

Query: 189 PSHKWNTKGK--LALVIGNEGKGISHNIKKQVDEMITIPMNGHVQSLNASVAAAILMYEV 246
               + N  G   LALVIG+EGKG+   +K++ D +I +PM G V SLNASVAA +LMYEV
Sbjct: 182 EDFR-NMDGNMPLALVIGSEGKGMGRLVKEKCDFLIKLPMAGKVTSLNASVAAGLLMYEV 240

Query: 247 FRNR 250
           +R R
Sbjct: 241 YRKR 244
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 555> which encodes the amino acid sequence <SEQ ID 556>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1037(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 206/248 (83%), Positives = 225/248 (90%), Gaps = 1/248 (0%)

Query: 3   MKDKQFKEESSDLVYGLHAVTESLRANTGNKLYLQDDLRGKNVDKVKALATEKKVSISWT 62
           M+DK    E++D+VYG+HAVTESL+ANTGNKLY+Q+DLRGK VD +K+LAT+KKV+ISWT
Sbjct: 10  MEDKD-TIETNDIVYGVHAVTESLQANTGNKLYIQEDLRGKKVDNIKSLATQKKVAISWT 68

Query: 63  PKKTLSDMTNGGVHQGFVLKVSEFAYADLSEIMTKAENEENPLILILDGLTDPHNLGSIL 122
           PKKTLS MT+G VHQGFVL+VS FAY D+ EI+  AE E NPLILILDGLTDPHNLGSIL
Sbjct: 69  PKKTLSQMTDGAVHQGFVLRVSAFAYTDVDEILEIAEQEANPLILILDGLTDPHNLGSIL 128

Query: 123 RTADATNVTGIIIPKHRSVGVTPVVSKTSTGAVEHVPIARVTNLSQTLDTLKDKEFWIFG 182
           RTADATNV G+IIPKHRSVGVTPVVSKTSTGAVEH+PIARVTNLSQTLD LK + FWIFG
Sbjct: 129 RTADATNVCGVIIPKHRSVGVTPVVSKTSTGAVEHIPIARVTNLSQTLDKLKARGFWIFG 188

Query: 183 TDMNGTPSHKWNTKGKLALVIGNEGKGISHNIKKQVDEMITIPMNGHVQSLNASVAAAIL 242
           TDMNGTPS  WNT GKLALVIGNEGKGIS NIKKQVDEMITIPMNGHVQSLNASVAAAIL
Sbjct: 189 TDMNGTPSDCWNTNGKLALVIGNEGKGISTNIKKQVDEMITIPMNGHVQSLNASVAAAIL 248

Query: 243 MYEVFRNR 250
           MYEVFRNR
Sbjct: 249 MYEVFRNR 256
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 167

A DNA sequence (GBSx0173) was identified in *S.agalactiae* <SEQ ID 557> which encodes the amino acid sequence <SEQ ID 558>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2187(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB11873 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 67/147 (45%), Positives = 94/147 (63%), Gaps = 2/147 (1%)

Query: 6   ILLVDGYNMIAFWKDTRQLFKSNRLEEAREVLLRKLNHYAHFEHIDIICVFDAQYVPGVR 65
           ILLVDGYNMI  W   + L K+N  EEAR+VL++K+  Y  +    +I VFDA  V G+
Sbjct: 3   ILLVDGYNMIGAWPQLKDL-KANSFEEARDVLIQKMAEYQSYTGNRVIVVFDAHLVKGLE 61

Query: 66  QRYDQYKISVIFTEEDETADSYIERAAAELNQSVLNLVSVATSDLNEQWTIFSQGALRVS 125
           ++   +++ VIFT+E+ETAD  IE+ A  LN ++   + VATSD  EQW IF QGALR S
Sbjct: 62  KKQTNHRVEVIFTKENETADERIEKLAQALN-NIATQIHVATSDYTEQWAIFGQGALRKS 120

Query: 126 ARELEQRVATVKSDLDKMSSQIDLSTP 152
           AREL + V T++  +++   +I    P
Sbjct: 121 ARELLREVETIERRIERRVRKITSEKP 147
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 559> which encodes the amino acid sequence <SEQ ID 560>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2465(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 130/167 (77%), Positives = 149/167 (88%), Gaps = 1/167 (0%)

Query: 3   KHSILLVDGYNMIAFWKDTRQLFKSNRLEEAREVLLRKLNHYAHFEHIDIICVFDAQYVP 62
           K  ILLVDGYNMIAFW+ TRQLFK+N+L++AR  LL KLNHYAHFE+I+IICVFDAQYVP
Sbjct: 2   KKRILLVDGYNMIAFWQSTRQLFKTNQLDQARNTLLTKLNHYAHFENINIICVFDAQYVP 61

Query: 63  GVRQRYDQYKISVIFTEEDETADSYIERAAAELNQSVLNLVSVATSDLNEQWTIFSQGAL 122
           G+RQRYDQY ISV+FTEEDETADSYIER AAELN + +++V VATSDLNEQWTIFSQGAL
Sbjct: 62  GLRQRYDQYYISVVFTEEDETADSYIERMAAELN-TAIHMVEVATSDLNEQWTIFSQGAL 120

Query: 123 RVSARELEQRVATVKSDLDKMSSQIDLSTPKLRPWNDEQLGKLKDFL 169
           RV+ARELEQRV TVK+DLDKMS  IDL TPKLRP++  QL +LKDF+
Sbjct: 121 RVTARELEQRVHTVKADLDKMSRDIDLKTPKLRPFDQGQLIQLKDFM 167
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 168

A DNA sequence (GBSx0174) was identified in *S.agalactiae* <SEQ ID 561> which encodes the amino acid sequence <SEQ ID 562>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4889(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB12951 GB:Z99109 yitS [Bacillus subtilis]
 Identities = 100/284 (35%), Positives = 157/284 (55%), Gaps = 6/284 (2%)

Query: 1   MTFKILTDSTSDLDEKWAQEHNVDIIGLTIELDGKTYETVGDEKITSDFLLERMQEGAKP 60
           MT  ++ DS +DL   + +E  +  I L + L  K +E      I +D + E MQ G  P
```

```
Sbjct: 1   MTVHLIADSATDLPRSYFEEKGIGFIPLRVSLGDKEFEDA--VTIHADQIFEAMQNGETP 58

Query: 61  TTSQINVGQFEEVFSTYAENDHALLYLALSSHLSGTYQSATIAREMVLDKYPDAQIEIVD 120
            TSQ +    + VF  YAE     LY+A SS LSGTYQ+A +    V +++PD  + ++D
Sbjct: 59  KTSQASPQTIKNVFLQYAETGDPALYIAFSSGLSGTYQTAVMIANEVKEEFPDFDLRVID 118

Query: 121 TMAASCGEGVLAMLATKERQEGKSLEEVKQKIESLLPKLNTYFLVDDLNHLMRSGRLSKG 180
            +  AS G G+    A      G +++E++  +++   +L   F VDDL +L R GR+SK
Sbjct: 119 SKCASLGYGLAVRHAADLCINGNTIQEIETSVKNFCSQLEHIFTVDDLTYLARGGRISKT 178

Query: 181 AAIIGSVAKIKPLLKLDSEGKLVPFAKTRGRKKGIK---EIVTQATKTLSYSTLIIAYSG 237
            +A +G +  IKPLL+++ +GKLVP  K RG+KK  K   E++ +     S  T+ I+Y+
Sbjct: 179 SAFVGGLLNIKPLLQME-DGKLVPLEKIRGQKKLFKRIIELMKERGDDWSNQTVGISYAA 237

Query: 238 EKDSAQVMKEQLLADERIEEVIIRPLGPVISAHVGSGALALFSL 281
            K+ A  MK  +    + +E+I+ P+   I +H G G LA+F L
Sbjct: 238 NKEKATDMKHLIEEAFKPKEIIMHPISSAIGSHAGPGTLAIFFL 281
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 563> which encodes the amino acid sequence <SEQ ID 564>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3247(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 167/286 (58%), Positives = 227/286 (78%)

Query: 1   MTFKILTDSTSDLDEKWAQEHNVDIIGLTIELDGKTYETVGDEKITSDFLLERMQEGAKP 60
            MTF I+TDST+DL++ WA++H++ +IGLTI  DG+ YETVG  +I+SD+LL++M+ G+ P
Sbjct: 1   MTFTIMTDSTADLNQTWAEDHDIVLIGLTILCDGEVYETVGPNRISSDYLLKKMKAGSHP 60

Query: 61  TTSQINVGQFEEVFSTYAENDHALLYLALSSHLSGTYQSATIAREMVLDKYPDAQIEIVD 120
            TSQINVG+FE+VF  +A N+ ALLYLA SS LSGTYQSA +AR++V + YPDA IEIVD
Sbjct: 61  QTSQINVGEFEKVFREHARNNKALLYLAFSSVLSGTYQSALMARDLVREDYPDAVIEIVD 120

Query: 121 TMAASCGEGVLAMLATKERQEGKSLEEVKQKIESLLPKLNTYFLVDDLNHLMRSGRLSKG 180
            T+AA+ GEG L +LA + R  GK+L E K  +E+++P+L TYFLVDDL HLMR GRLSKG
Sbjct: 121 TLAAAGGEGYLTILAAEARDSGKNLLETKDIVEAVIPRLRTYFLVDDLFHLMRGGRLSKG 180

Query: 181 AAIIGSVAKIKPLLKLDSEGKLVPFAKTRGRKKGIKEIVTQATKTLSYSTLIIAYSGEKD 240
            +A +GS+A IKPLL +D EGKLVP AK RGR+K IKE+V Q  K ++ ST+I++Y+ ++
Sbjct: 181 SAFLGSLASIKPLLWIDEEGKLVPIAKIRGRQKAIKEMVAQVEKDIADSTVIVSYTSDQG 240

Query: 241 SAQVMKEQLLADERIEEVIIRPLGPVISAHVGSGALALFSLGEENR 286
            SA+ ++E+LLA E I +V++ PLGPVISAHVG   LA+F +G+ +R
Sbjct: 241 SAEKLREELLAHENISDVLMMPLGPVISAHVGPNTLAVFVIGQNSR 286
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 169

A DNA sequence (GBSx0175) was identified in *S.agalactiae* <SEQ ID 565> which encodes the amino acid sequence <SEQ ID 566>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.76   Transmembrane   43 -  59 (  40 -  62)
```

```
----- Final Results -----
               bacterial membrane --- Certainty=0.4503(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 170

A DNA sequence (GBSx0176) was identified in *S.agalactiae* <SEQ ID 567> which encodes the amino acid sequence <SEQ ID 568>. This protein is predicted to be ribosomal protein L13 (rplM). Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3426(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9507> which encodes amino acid sequence <SEQ ID 9508> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB03887 GB:AP001507 ribosomal protein L13 [Bacillus halodurans]
 Identities = 89/144 (61%), Positives = 113/144 (77%)

Query: 36  KTTFMAKPGQVERKWYVVDAADVPLGRLSAVVASVLRGKNKPTFTPHTDTGDFVIVINAE 95
           +TT+MAKP +VERKWYVVDA    LGRL++ VAS+LRGK+KPT+TPH DTGD VI+INAE
Sbjct: 2   RTTYMAKPNEVERKWYVVDAEGQTLGRLASEVASILRGKHKPTYTPHVDTGDHVIIINAE 61

Query: 96  KVKLTGKKASDKIYYTHSMYPGGLKQISAGELRSKNAVRLIEKSVKGMLPHNTLGRAQGM 155
           K+ LTG K  DKIYY HS +PGGLK+  A ++R+     +++E ++KGMLP NTLGR QGM
Sbjct: 62  KIHLTGNKLQDKIYYRHSGHPGGLKETRAADMRANKPEKMLELAIKGMLPKNTLGRKQGM 121

Query: 156 KLKVFVGGEHTHAAQQPEVLDISG 179
           KL V+ G EH H AQ+PEV ++ G
Sbjct: 122 KLHVYAGSEHKHQAQKPEVYELRG 145
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 569> which encodes the amino acid sequence <SEQ ID 570>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4249(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 167/184 (90%), Positives = 171/184 (92%), Gaps = 4/184 (2%)

Query: 1   MFTPFVRPRNLSNTLVDRNIHT--CKQ-KRIRIGEIMNKTTFMAKPGQVERKWYVVDAAD 57
           +FTPF RPRNL NT  D   H   CKQ  RIRIGEIMNKTTFMAKPGQVERKWYVVDAAD
```

```
Sbjct: 1   LFTPFERPRNLPNTF-DGTEHPSPCKQILRIRIGEIMNKTTFMAKPGQVERKWYVVDAAD 59

Query: 58  VPLGRLSAVVASVLRGKNKPTFTPHTDTGDFVIVINAEKVKLTGKKASDKIYYTHSMYPG 117
           VPLGRLSAVVASVLRGKNKPTFTPHTDTGDFVIVINAEKVKLTGKKA+DK+YYTHSMYPG
Sbjct: 60  VPLGRLSAVVASVLRGKNKPTFTPHTDTGDFVIVINAEKVKLTGKKATDKVYYTHSMYPG 119

Query: 118 GLKQISAGELRSKNAVRLIEKSVKGMLPHNTLGRAQGMKLKVFVGGEHTHAAQQPEVLDI 177
           GLK I+AGELRSKNAVRLIEKSVKGMLPHNTLGRAQGMKLKVFVGGEHTHAAQQPEVLDI
Sbjct: 120 GLKSITAGELRSKNAVRLIEKSVKGMLPHNTLGRAQGMKLKVFVGGEHTHAAQQPEVLDI 179

Query: 178 SGLI 181
           SGLI
Sbjct: 180 SGLI 183
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 171

A DNA sequence (GBSx0177) was identified in *S.agalactiae* <SEQ ID 571> which encodes the amino acid sequence <SEQ ID 572>. This protein is predicted to be 30S ribosomal protein S9 (rpsI). Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1761(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB11926 GB:Z99104 ribosomal protein S9 [Bacillus subtilis]
 Identities = 88/130 (67%), Positives = 105/130 (80%)

Query: 1   MAQAQYAGTGRRKNAVARVRLVPGTGKITINKKDVEEYIPHADLRLVINQPFAVTSTQGS 60
           MAQ QY GTGRRK++VARVRLVPG G+I +N +++ E+IP A L   I QP  +T T G+
Sbjct: 1   MAQVQYYGTGRRKSSVARVRLVPGEGRIVVNNREISEHIPSAALIEDIKQPLTLTETAGT 60

Query: 61  YDVFVNVVGGGYAGQSGAIRHGISRALLEVDPDFRDSLKRAGLLTRDARMVERKKPGLKK 120
           YDV VNV GGG +GQ+GAIRHGI+RALLE DP++R +LKRAGLLTRDARM ERKK GLK
Sbjct: 61  YDVLVNVHGGGLSGQAGAIRHGIARALLEADPEYRTTLKRAGLLTRDARMKERKKYGLKG 120

Query: 121 ARKASQFSKR 130
           AR+A QFSKR
Sbjct: 121 ARRAPQFSKR 130
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 573> which encodes the amino acid sequence <SEQ ID 574>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1865(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 124/130 (95%), Positives = 129/130 (98%)

Query: 1   MAQAQYAGTGRRKNAVARVRLVPGTGKITINKKDVEEYIPHADLRLVINQPFAVTSTQGS 60
```

```
            MAQAQYAGTGRRKNAVARVRLVPGTGKIT+NKKDVEEYIPHADLRL+INQPFAVTST+GS
Sbjct: 1    MAQAQYAGTGRRKNAVARVRLVPGTGKITVNKKDVEEYIPHADLRLIINQPFAVTSTEGS 60

Query: 61   YDVFVNVVGGGYAGQSGAIRHGISRALLEVDPDFRDSLKRAGLLTRDARMVERKKPGLKK 120
            YDVFVNVVGGGY GQSGAIRHGI+RALL+VDPDFRDSLKRAGLLTRDARMVERKKPGLKK
Sbjct: 61   YDVFVNVVGGGYGGQSGAIRHGIARALLQVDPDFRDSLKRAGLLTRDARMVERKKPGLKK 120

Query: 121 ARKASQFSKR 130
            ARKASQFSKR
Sbjct: 121 ARKASQFSKR 130
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 172

A DNA sequence (GBSx0178) was identified in *S.agalactiae* <SEQ ID 575> which encodes the amino acid sequence <SEQ ID 576>. This protein is predicted to be recombinase (b1345). Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1939(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAG29618 GB:AF217235 integrase-like protein [Staphylococcus
            aureus]
 Identities = 127/386 (32%), Positives = 205/386 (52%), Gaps = 18/386 (4%)

Query: 3    IHKYPSKKAKNGYLYFVKIYMVKD---SQRADHIKRGFRTRKEAKDYEARLIYLKASGKL 59
             I KY  K     Y++    Y+  D    ++    +RGF+T +EAK  EA+L   +
Sbjct: 2    IKKYKKKDGSTAYMFVA--YLGTDPITGKQKRTTRRGFKTEREAKIAEAKL---QTEVSQ 56

Query: 60   EEFIKPTHKTYNEIFEKWYQAYQDMVEPTTASRTLDMFRLHILPVMGDLPISKISPLDCQ 119
               F+      T+ E++E W + YQ+ V  +T  R L +F   IL    D+PI KI+   CQ
Sbjct: 57   NGFLNNDITTFKEVYELWLEQYQNTVRESTYQRVLTLFDTAILEHFQDVPIKKITVPYCQ 116

Query: 120  NFITDKAKTFKNIKQIKSYTGKVFDFAIKMKLLKHNPMAEIIMPKRKKTRIE---NYWTV 176
               I    K + +IK I+ YT  VF +A+ +K++  NP A    P++K+ + +    Y++
Sbjct: 117  KVINKWNKKYSDIKAIRIYTSNVFKYAVSLKIIVDNPFAHTKAPRKKEAQQDASTKYYSS 176

Query: 177  QELQEFLAIVLQEEPYKHYALFRLLAYSGLRKGELYALKWADIDFQTETLSVDKSLGR-L 235
              EL++FL  V  E+   +YA+FR LA++G R+GEL AL W DIDF  +T+S++K+  R
Sbjct: 177  DELKQFLTFV--EDDPLYYAIFRTLAFTGFRRGELMALTWNDIDFTKQTISINKTCARGA 234

Query: 236  DGQAIEKGTKNDFSVRKIKLDSETISILQEWKSISQKEKAQLAVAPLSIEQDFLFTYCTR 295
             + + + +  K   S R I +D +T S+L+ W++  + E  +      S +   +FT
Sbjct: 235  NYKLVIQEPKTKSSHRTISIDDKTASVLKSWRTHQRVESLKYG-HNTSDKHQHVFTTVRD 293

Query: 296  SGSIEPLHADYINNVLSRIIRKHGLKKISPHGFRHTHATLMIEIGVDPVNTAKRLGHASS 355
             +   +PL+ ++ N  L  I  K+  K+I  HGFRHTH +L+ E G+       RLGH
Sbjct: 294  N---KPLYPEHCNKALDLICEKNSFKRIKVHGFRHTHCSLLFEAGLSIQEVQDRLGHGDI 350

Query: 356  QMTLDTYSHSTTTGEDRSVKQFADYL 381
             + T+D Y+H T    D+   +FA Y+
Sbjct: 351  KTTMDIYAHVTEKQRDQVADKFAKYI 376
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 577> which encodes the amino acid sequence <SEQ ID 578>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3445(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 109/386 (28%), Positives = 185/386 (47%), Gaps = 28/386 (7%)

Query: 3   IHKYPSKKAKNGYL-YFVKIYMVKDSQRADHIKRGF--RTRKEA--KDYEARLIYLKASG 57
           I K    K KNG + Y   IY+  D      +K     RTRKE   K   A+  +L
Sbjct: 6   IMKITEHKKKNGTIVYRASIYLGIDQMTGKRVKTSITGRTRKEVNQKAKHAQFDFLSNGS 65

Query: 58  KLEEFIKPTHKTYNEIFEKWYQAYQDMVEPTTASRTLDMFRLHILPVMGDLPISKISPLD 117
            ++   K   KT+ E+   W + Y+  V+P T   T+     HI+P +G++ + KI+  D
Sbjct: 66  TIKR--KVVIKTFKELSHLWLETYKLTVKPQTYDATVTRLNRHIMPTLGNMKVDKITASD 123

Query: 118 CQNFITDKAKTFKNIKQIKSYTGKVFDFAIKMKLLKHNPMAEIIMPKRK---KTRIENYW 174
            Q  I   +K + N   ++S   KV    + + L+ +N   +II+P+++   K +++ +
Sbjct: 124 IQMLINRLSKYYVNYTAVRSVIRKVLQQGVLLGLIDYNSARDIILPRKQPNAKKKVK-FI 182

Query: 175 TVQELQEFLAIVLQEEPYKHY------ALFRLLAYSGLRKGELYALKWADIDFQTETLSV 228
              +L+ FL   L+   +K Y       L++LL  +GLR GE  AL+W DID +  T+++
Sbjct: 183 DPSDLKSFLE-HLETSQHKRYNLYFDAVLYQLLLSTGLRIGEACALEWGDIDLENGTIAI 241

Query: 229 DKSLGRLDGQAIEKGTKNDFSVRKIKLDSETISILQEWKSISQKEKAQLAVAPLSIEQDF 288
           +K+  +          K     R I +D +T+  L+    + Q  + QL     +   +
Sbjct: 242 NKTYNK--NLKFLSTAKTQSGNRVISVDKKTLRSLK----LYQMRQRQLFNEVGARVSEV 295

Query: 289 LFTYCTRSGSIEPLHADYINNVLSRIIRKHGLKKISPHGFRHTHATLMIEIGVDPVNTAK 348
           +F   TR    +  +A    + L    ++ G+++ + H FRHTHA+L++  G+
Sbjct: 296 VFATPTR----KYFNASVRQSALDTRCKEAGIERFTFHAFRHTHASLLLNAGISYKELQY 351

Query: 349 RLGHASSQMTLDTYSHSTTTGEDRSV 374
           RLGHA+  MTLDTY H +   E  +V
Sbjct: 352 RLGHANISMTLDTYGHLSKGKEKEAV 377
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 173

A DNA sequence (GBSx0179) was identified in *S.agalactiae* <SEQ ID 579> which encodes the amino acid sequence <SEQ ID 580>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2477(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF63067 GB:AF158600 putative DNA binding protein
          [Streptococcus thermophilus bacteriophage Sfi11]
 Identities = 32/70 (45%), Positives = 46/70 (65%), Gaps = 3/70 (4%)

Query: 3  NRLKELRKDKGLTQADLAKVINTNQSQYGKYENGKTSLSIENSKILADFFGVSIPYLLGL 62
           NRL  LR+ + +T+ +LA+ I  ++     K E+G + +S   +K LADFFGVS+ YLLGL
Sbjct: 2  NRLYLLRESRKITRVELAEKIGVSKLTVLKLEHGTSKISRREAKKLADFFGVSVGYLLGL 61
```